CRYSTALLOGRAPHIC STUDIES OF *RHIZOMUCOR MIEHEI* ASPARTIC PROTEINASE AND ITS PEPSTATIN A COMPLEX

A Thesis Submitted to the College of

Graduate Studies and Research in

Partial Fulfillment of Requirements

for the Degree of Doctor of Philosophy

in the Department of Chemistry

University of Saskatchewan

Saskatoon

By

Jian Yang

Spring, 1998

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any matter, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make use of materials in the thesis in whole or part should be addressed to:

Head of the Department of Chemistry
University of Saskatchewan
110 Science Place
Saskatoon, Saskatchewan
S7N 5C9 CANADA

ABSTRACT

The crystal structures of *Rhizomucor miehei* aspartic proteinase (RMP) and its pepstatin A complex have been determined at 2.15 Å and 2.7 Å, respectively. The structure of the native enzyme was refined to a crystallographic *R*-factor of 21.5% (*R*-free = 28.1%). RMP contains two domains that consist predominantly of β-sheets. A large substrate binding cleft is clearly visible between the two domains, and the two catalytic residues Asp38 and Asp237 are located in the middle of the cleft with a water molecule bridging the carboxyl groups of Asp38 and Asp237. It is proposed that the optimal pH of each aspartic proteinase is determined by the electrostatic potential at the active site, which, in turn, is determined by the positions and orientations of all the residues near the active site. RMP is the most glycosylated and most thermally stable among the aspartic proteinases. It is proposed that the highly flexible carbohydrates act as heat reservoirs to stabilize the conformation of RMP and thereby provide the enzyme with high thermal stability. Three-dimensional structural and sequence alignments of RMP with other aspartic proteinases suggest that RMP and *Mucor pusillus* aspartic proteinase (MPP) diverged from the main stream of aspartic proteinases at an early stage of evolution.

The structure of the RMP-pepstatin A complex was refined to a crystallographic *R*-value of 19.3% (*R*-free = 28.0%). In the final model, a pepstatin A molecule fits into the large substrate-binding cleft between the two domains of RMP in an extended conformation up to the alanine residue at the P2' position. The statine residue at the "P3'-P4'" position forms an inverse γ-turn (P3'–P1'), with its leucyl side chain binding into the S1' subsite. The inhibitor interacts with the residues of the substrate binding pocket by either hydrogen bonds or hydrophobic interactions, or both.

Acknowledgments

It is my pleasure to express my gratitude to my research supervisor, Dr. J. Wilson Quail, for his constant assistance, considerable patience and great guidance during my tenure as a graduate student. I cherish his friendship and his encouragement to develop myself as an independent investigator. I owe him my sincerest thanks for having been given the opportunity to work on such an interesting and important project in his laboratory.

I would like to thank Dr. Louis T.J. Delbaere for his constructive discussions, valuable suggestions, encouragement and friendship. Talking to him is always an enjoyable thing.

I would also like to extend sincere thanks to past and present members of my supervisory committee: Drs. J.R. Dimmock, V.S. Gupta, D. Russell, G. Tourigny and R.J. Woods for their valuable comments and assistance during this project. Special thanks go to Dr. Dimmock for his financial support during the first year of my study at the University of Saskatchewan. Our collaboration with Dr. Dimmock's laboratory has always been enjoyable.

I value the friendship and cooperative spirit that exists in this laboratory. Several present and past members of this laboratory deserve acknowledgments. Ms. Margaret Vandonselaar and Dr. Zongchao Jia played an important role thrcugh their work on establishing the initial crystallization conditions and preliminary X-ray studies of RMP. Without their work, my task would have been much more difficult, if not impossible. Dr. Lata Prasad has helped me in many aspects, from the crystallography theories to the computing programs, from the experiment setups to the discussion of the results. She is always willing to help. She deserves my sincere thanks. Ms. Umarani Pugazhenthi gave so much help when I joined the lab. She taught me how to grow a crystal, mount a crystal, take precession photographs, collect data and solve structures on computers. I really appreciate her friendship. I wish to thank Drs. Allan Matte and Leslie Tari for their help with the computing programs and for their helpful comments and discussions. In addition, a word of appreciation should be addressed to other members in the laboratory, Scott Napper and Gerald Audette, for making the laboratory such a pleasant place to work in.

I gratefully acknowledge the Department of Chemistry, and the College of Graduate Studies and Research, University of Saskatchewan for awarding

me a graduate scholarship (1993-1997). This research project is supported by a research grant (to J.W. Quail) from the Natural Sciences and Engineering Research Council of Canada. The University of Saskatchewan President's Fund should also be acknowledged for awarding me travel support in three consecutive years (1995-1997).

My sincere thanks also go to Dr. P. Schneider, Novo Nordisk, Denmark, for providing the purified RMP enzyme, Drs. A. Teplyakov and Z. Dauter, EMBL, Germany, for their help with the synchrotron data collection of the native RMP enzyme, and Dr. R.M. Sweet, Brookhaven National Laboratory, USA, for his help with the synchrotron data collection of the RMP-pepstatin A complex.

Most importantly, I am grateful to my parents for their support through many years of education. Without their encouragement, I would not have been able to approach to this step.

DEDICATION

This thesis is dedicated to my parents, Yulan Gao and Jinting Yang, who provided me with so much to make my dreams a reality.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

CCD	charge coupled device
CCP4	Collaborative Computational Project #4
EMBL	European Molecular Biology Laboratory
Gdn	guanidine
IVA	isovaleric acid
MAD	multiple wavelength anomalous diffraction
MIR	multiple isomorphous replacement
MPD	2-methyl-2,4-pentanediol
MPP	*Mucor pusillus* aspartic proteinase
MR	molecular replacement
Nle	norleucine
PEG	polyethylene glycol
RMP	*Rhizomucor miehei* aspartic proteinase
STA	statine
V_m	Mathews coefficient

Chapter 1. Introduction

1.1 An overview

Aspartic proteinases (EC 3.4.23) are a group of proteolytic enzymes, which are usually optimally active at acidic pH (except renins). They are so called because of the presence of two catalytically-essential aspartic acid residues at the active site. Aspartic proteinases, together with serine, cysteine and metalloproteinases, form four major classes of proteolytic enzymes. Aspartic proteinases are widely distributed, having been found in micro-organisms, plants and mammals. Because they cleave peptide bonds, aspartic proteinases play important roles in many biological processes. These enzymes are usually inhibited by pepstatin, a pentapeptide produced by *Streptomyces* (Umezawa *et al.*, 1970). The chemical structure of the pepstatin molecule is shown in Figure 1.1. It contains two dipeptide analogue statine (STA) residues and one isovaleric acid (IVA) residue. This molecule is usually called pepstatin A, while its derivatives formed by replacing isovaleryl with n-caproyl and iso-caproyl are called pepstatin B and pepstatin C. The aspartic proteinases are broadly divided into two main groups: pepsin-like and retroviral enzymes. The pepsin-like enzymes can be further divided into the

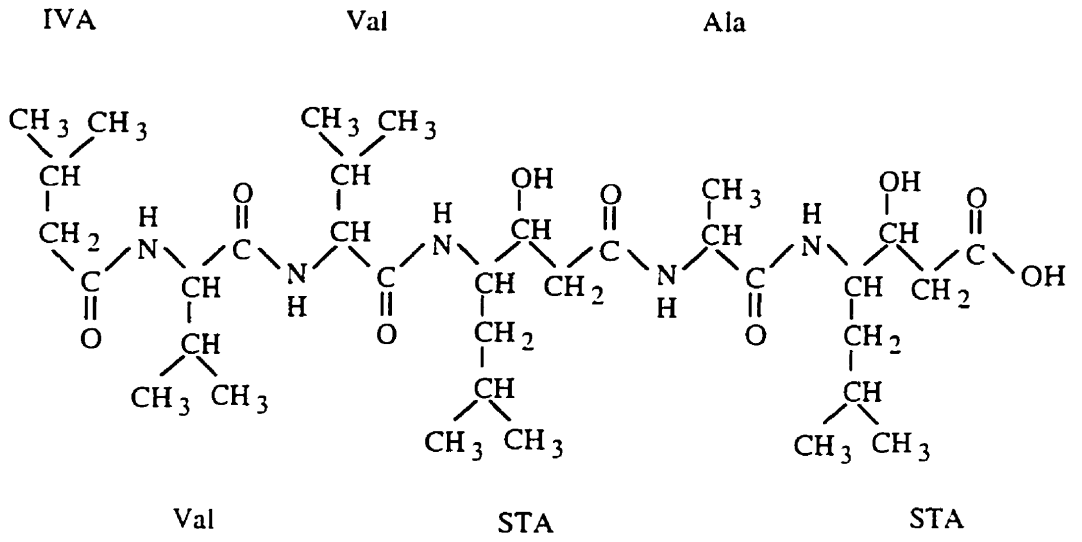


Figure 1.1. Chemical structure of pepstatin (Umezawa *et al.*, 1970). IVA represents isovaleric acid and STA represents statine.

following groups: pepsins and chymosins, cathepsins, renins and fungal aspartic proteinases (Szecsi, 1992). The pepsin-like aspartic proteinases are monomers and have molecular weights of approximately 35 ~ 38 kDa. These enzymes contain between 320 and 361 amino acid residues, and consist of two domains of approximately the same size. The two domains are related by an approximate 2-fold symmetry (Tang *et al.*, 1978). The two catalytic aspartate residues are contributed by each of the two domains, respectively. The retroviral aspartic proteinases have between 99 and 150 amino acid

residues, existing as homodimers in the active form. With the discovery that HIV proteinase is a retroviral aspartic proteinase, aspartic proteinase studies have become a very important area in the structure-based drug/inhibitor design research.

Because of their digestive characteristics, many aspartic proteinases have been used as milk-clotting enzymes in cheese industry. These enzymes can cleave κ-casein in milk into two parts: a para-κ- portion and a macropeptide. This cleavage disturbs the stability of the casein micelles and causes the aggregation of the proteins, forming a curd. The most prominent milk-clotting enzyme is chymosin, which is usually obtained from the stomach of young calves. Chymosin, a pepsin-like aspartic proteinase, has been used in cheese making since antiquity. Two types of chymosins, A and B, are present naturally. These two types of chymosins differ by a single amino acid residue at position 244 in their amino acid sequences. Residue 244 is an aspartate in chymosin A and a glycine in chymosin B. Chymosin possesses high milk-clotting activity due to its specific cleavage of κ-casein (Brown & Ernstrom, 1988). Chymosin is the enzyme of choice and the standard against which all the other aspartic proteinases are measured. However, due to a worldwide shortage of calf chymosin, attempts are being made to find substitutes for this enzyme. One of the substitutes is *Rhizomucor miehei*

aspartic proteinase (RMP, EC 3.4.23.23). It exhibits high milk-clotting activity along with low proteolytic activity compared with other aspartic proteinases (Yada & Nakai, 1986a,b). RMP has been used to make excellent quality Cheddar cheese (Prins & Nielsen, 1970). However, it is the most thermally stable enzyme among the aspartic proteinases (Hyslop *et al.*, 1979), and therefore is of limited use as a milk coagulant. The high thermal stability results in the survival of the enzyme activity after cooking of the curd and thus may cause ill flavor in cheese during long maturating periods (Yamashita *et al.*, 1994). The X-ray crystallographic study of RMP was carried out in order to determine the reason for its high thermal stability and provide a guideline for further modification of this enzyme to reduce its thermal stability in order to try to make use of its milk-clotting activity.

1.2 The making of cheese

Cheese making is perhaps one of the oldest means of preserving food. Cheese is the generic name of a group of products cultured/fermented from milk. The first cheese was believed to be made in the Tigris-Euphrates Valley about 8000 years ago in what is now Iraq (Holsinger *et al.*, 1995). Nowadays approximately 4000 different cheese varieties, with a variety of flavor and/or texture characteristics, are produced worldwide (Steele, 1995). The ability to manufacture such a variety of products from milk is due to the different conditions and microorganisms used in the manufacturing.

1.2.1 The composition of milk

Milk is the liquid food secreted by all species of mammals to supply their young with nutrition and immunological protection. The composition of milk has been reviewed by Johnson (1974) and Jenness (1988). It is an extremely complex mixture of components, consisting of water, lipids, carbohydrates, proteins, salts, and many miscellaneous compounds. It contains probably about 10^5 different kinds of molecules.

The largest component in milk is water, which can be up to 87% (Johnson, 1974). All the other components in milk are either dissolved or suspended in water. A small percentage of the water is hydrated to lactose and salts, or is bound to proteins. The lipids, which are usually called "fat", are 98% triglycerides, mostly as globules with diameters of 0.1 to 1.5 μm. Carbohydrate is present mainly as lactose. There are several classes of proteins. The largest group (approximate 80%) are phosphoproteins called "caseins". They are associated with calcium in the casein micelles, which are 20-300 μm in diameter (Jenness, 1988). The caseins can be further divided into four major groups: α_{s1}-, α_{s2}-, β-, and κ-caseins. The other milk proteins are called "whey proteins" and include β-lactoglobulin, α-lactalbumin, blood serum albumin and immunoglobulins. The ions are mainly Na^+, K^+, Ca^{2+}, Mg^{2+}, PO_4^{3-}, Cl^-, CO_3^{2-}, SO_4^{2-} and citrate anion. Besides these compounds, milk also contains trace elements, such as aluminum, bromine, copper, zinc, iron, manganese, and silicon.

1.2.2 The procedures involved in cheese making

The cheese making process can be divided into manufacturing and ripening phases. The manufacturing stage refers to those events that occur

during the first 24 hours (Fox, 1993). The basic steps in the manufacturing stage are: acidification; coagulation; dehydration; shaping; and salting (Holsinger *et al.*, 1995). During the ripening stage, the curd undergoes significant biochemical changes which are regulated by the moisture content, the salt content and the microflora. The differentiation in flavor, odor and texture of various cheeses is determined by the biochemical changes during this stage.

Manufacturing stage

Acidification

Acidification of cheese milk is the first step in cheese making. It is a progressive development of acidity throughout the manufacturing stage. Acid production is very important to the quality of cheese. Acidification affects the milk coagulant's activity, curd strength, syneresis, pH and growth of non-starter microorganisms. If the milk is too acidic, the cheese will be crumbly; if too basic, the cheese is pasty and sticky.

Coagulation

Coagulation is the second and the most important step in the manufacturing stage. The curd is formed in this stage in one of two ways: acid or enzymatic coagulation. In acid-curd cheeses (cottage, cream), the curd is formed by the direct addition of acid to the milk or by lactic acid produced by the fermentation of lactose. Nowadays, coagulation by various aspartic proteinases has become predominant in cheese-making. Of the many aspartic proteinases obtained from animal, plant and microbial sources that have been used for the manufacture of cheese and other foods since antiquity, the most common one is chymosin. The properties of curds formed by this method are quite different from those of the acid precipitated curds. These curds have better syneresis properties which make it possible to produce low moisture curd without hardening.

Dehydration

From this point, the post-coagulation changes start. These changes will determine the quality of the final products. The curd synereses quickly when cut, expelling the whey, although it is quite stable if it is quiescent. The syneresis is controlled by quite a few factors: milk composition, Ca^{2+}

concentration, pH of the whey, cooking temperature *etc.* The extent of dehydration in this step determines the moisture content of the curd, which will affect the odor, flavor and quality of the final cheese products. It is from this step that the differentiation of cheeses begins.

Salting

Almost all cheese varieties are salted. Salt (NaCl) is added to control the water activity, microbial growth and activity, enzyme activity and the physical changes in the cheese proteins during the ripening stage. Salting affects the flavor and the quality of cheeses. Usually, salting is done by immersing the cheese blocks in a brine solution (20–22% NaCl + 0.3% $CaCl_2$) at 12°C, from which the salt diffuses into the cheese (Rosenthal, 1991).

Ripening stage

After the manufacturing stage, the cheeses are stored in a ripening room from two weeks to several months. The cheese ripening involves several biochemical changes. These changes include proteolysis, glycolysis and lipolysis. Proteolysis is the most significant process because it results in

the breakdown of the structural network of casein formed during the manufacturing stage and the formation of a soft and smooth mature cheese. Flavor compounds produced by proteolysis are released during the ripening stage. The major proteolytic changes during maturation of cheese are the conversion of protein to large peptides, large peptides to small peptides, and small peptides to free amino acids, amides and ammonia (Farkye, 1995). The rate and extent of proteolysis in the ripening stage are determined by the types and activities of the milk-clotting enzymes, pH value, moisture content and salt, chamber temperature and surface treatment during ripening (Rosenthal, 1991). Usually less than 10% of the milk-clotting activity used for cheese making is retained in cheese. The reduction in milk-clotting activity of the coagulant enzymes is usually done by heating during the manufacturing stage.

1.2.3 κ–casein

The caseins were defined originally as a group of phosphoproteins which precipitate from raw skim milk upon acidification to pH 4.6 at 20°C (Whitney *et al.*, 1976). They make up to 80% of the total proteins in milk. With the development of the primary structural studies, caseins are now

classified according to their chemical structures. Caseins can be divided into four major groups, α_{s1}-, α_{s2}-, β-, and κ- in the ratio 3:0.8:3:1. These casein proteins are associated with calcium in "casein micelles", which are 20-300 μm in diameter (Jenness, 1988). The outside surface of micelles contain a high concentration of κ-caseins (Walstra & Jenness, 1984).

The κ-caseins are soluble over a very broad range of calcium ion concentrations (Waugh & Von Hippel, 1956). This solubility stabilizes the casein micelles (Waugh & Von Hippel, 1956). The κ-caseins are the only major component of the casein complex which contains carbohydrate and occur as a mixture of polymers held together by disulfide bonds. With the addition of reducing agents, like mercaptoethanol and dithiothreitol, the polymers can be converted completely to monomers. The relatively high heterogeneity of caseins is due to the genetic differences, variation in carbohydrate content and/or phosphate content and the para-κ-portion of the molecule. Two genetic variants of the κ-caseins are known: A and B (Thompson, 1970; Mackinlay & Wake, 1971). Both of these two variants show multiple bands in the alkaline urea gel electrophoresis in the presence of mercaptoethanol, with the A variant bands migrating faster (Swaisgood,

1975). The A variant is predominant in most species of domesticated animals (Aschaffenburg, 1968).

The amino acid sequence of the two κ-casein variants have been determined (Jollès *et al.*, 1972a, b; Mercier *et al.*, 1973) (Figure 1.2). Both variants contain 169 amino acid residues. The sequence difference between these two variants is located at residues 136 and 148. In variant A, residues 136 and 148 are threonine and aspartic acid, respectively. In variant B, these two residues are isoleucine and alanine, respectively. Among all the caseins, the κ-caseins are the ones that are most readily cleaved by chymosin (Kalan & Woychik, 1965). The κ-caseins are split by chymosin at the Phe105-Met106 bond, producing the cationic para-κ-portion of the molecule (residues 1-105) and the anionic macropeptide (residues 106-169). Splitting of κ-caseins destroys the stability of the milk system and the caseins begin to aggregate, forming a curd. Further proteolysis of κ-caseins during the ripening will determine the flavor, body and texture development in cheese.

```
QEQNQEQPIR CEKDERFFSD KIAKYIPIQY VLSRYPSYGL NYYQQKPVAL  50

INNQFLPYPY YAKPAAVRSP AQILQWQVLS NTVPAKSCQA QPTTMARHPH 100

   105↓106
PHLSFMAIPP KKNQDKTEIP TINTIASGEP TSTPTTEAVE STVATLEDSP 150
                                        I (variant B)  A

EVIESPPEIN TVQVTSTAV 169
```

Figure 1.2. Amino acid sequence of variant A of κ-casein in the one-letter code for amino acids (IUPAC-IUB Commission on Biochemical Nomenclature, 1972); variant B sequence differences are also denoted.

1.2.4 Milk-clotting enzymes

One of the most important steps in cheese making is clotting the milk. Although the curd can be formed in several ways, clotting by milk coagulant enzymes has become predominant in the modern cheese industry. Many proteolytic enzymes will clot milk and milk-clotting enzymes have been obtained from all kinds of sources, such as animals, plants, fungi and bacteria. To be a good milk-clotting enzyme, the enzyme should possess the following characteristics:

(1) High milk-clotting activity due to selective cleavage of κ-casein.

(2) Low non-specific proteolytic activity.

(3) Low thermal stability.

Milk-clotting enzymes from various sources have been made commercially available. Examples of these enzymes are chymosin, porcine pepsin, bovine pepsin and chicken pepsin from animals; endothiapepsin, *Rhizomucor miehei* aspartic proteinase (RMP) and *Mucor pusillus* aspartic proteinase (MPP) from fungi; and papain, chymopapain, ficin and bromelain from plants. Among these enzymes, calf chymosin is still the ideal enzyme and the standard against which other milk-clotting enzymes are compared. Because of its relatively high milk-clotting activity, RMP has been used as a substitute for calf chymosin and has about a 30% share of the world milk-clotting enzyme market.

1.3 *Rhizomucor miehei* aspartic proteinase

Rhizomucor miehei aspartic proteinase (RMP) is the native enzyme form obtained by fermentation of *Rhizomucor miehei*. RMP was extensively studied in the 1970s and 1980s by several groups, Sternberg, Rickert and colleagues, and Foltmann and colleagues. RMP is identical to the enzyme which is commercially available as Rennilase (Bech & Foltmann, 1981; Boel *et al.*, 1986). The enzyme contains 361 amino acid residues and the molecular weight is 38.7 kilodaltons. RMP is synthesized as a zymogen containing a 47-amino-acid propeptide from a precursor containing 430 amino acid residues (Bech & Foltmann, 1981; Boel *et al.*, 1986; Gray *et al.*, 1986). Like many other aspartic proteinases, RMP can also be inhibited by pepstatin A, the common aspartic proteinase inhibitor.

1.3.1 The amino acid sequence of RMP

The amino acid sequence of RMP was identified from the sequence determination of its zymogen (Bech & Foltmann, 1981; Boel *et al.*, 1986; Gray *et al.*, 1986). The amino acid sequence of RMP is shown in Figure 1.3. The

enzyme is glycosylated at Asn79 and Asn188 (Boel *et al.*, 1986). Two disulfide bonds are present in the molecule, Cys51-Cys57 and Cys272-Cys316 (Boel *et al.*, 1986).

```
AAADGSVDTP GYYDFDLEEY AIPVSIGTPG QDFLLLFDTG SSDTWVPHKG  50
CTKSEGCVGS RFFDPSASST FKATNYNLNI TYGTGGANGL YFEDSIAIGD 100
ITVTKQILAY VDNVRGPTAE QSPNADIFLD GLFGAAYPDN TAMEAEYGST 150
YNTVHVNLYK QGLISSPLFS VYMNTNSGTG EVVFGGVNNT LLGGDIAYTD 200
VMSRYGGYYF WDAPVTGITV DGSAAVRFSR PQAFTIDTGT NFFIMPSSAA 250
SKIVKAALPD ATETQQGWVV PCASYQNSKS TISIVMQKSG SSSDTIEISV 300
PVSKMLLPVD QSNETCMFII LPDGGNQYIV GNLFLRFFVN VYDFGNNRIG 350
FAPLASAYEN E 361
```

Figure 1.3 The amino acid sequence of RMP using the one-letter code for amino acid residues (IUPAC-IUB Commission on Biochemical Nomenclature, 1972).

1.3.2 The milk-clotting activity of RMP

RMP was first reported by Prins and Nielsen (1970) to be used as a milk coagulant in making good quality Cheddar cheese. It is quite stable over a broad range of conditions, with the maximum stability between pH 4.0 and 6.0 (Sternberg, 1971). Bond specificity studies with synthetic substrates

(Sternberg, 1972) show that RMP cleaves peptide bonds which have an aromatic amino acid on the carbonyl side of the peptide bond. Like other aspartic proteinases, RMP cleaves κ-casein at the peptide bond between Phe105-Met106. The kinetic study results of RMP and several other aspartic proteinases catalyzing the cleavage of a synthetic hexapeptide and κ-casein (Martin *et al.*, 1980) show that RMP exhibits maximum activity at a temperature above 63°C and pH around 4.7, with the synthetic hexapeptide Leu-Ser-Phe(NO_2)-Nle-Ala-Leu-OMe as the substrate. Nle in this hexapeptide sequence refers to norleucine. The site cleaved by RMP in this substrate is the peptide bond between Phe(NO_2) and Nle. Under the same experiment conditions the k_{cat}/K_m for RMP (41.5 $mM^{-1}s^{-1}$) is higher than that for chymosin (25 to 30 $mM^{-1}s^{-1}$). This indicates that this hexapeptide is a slightly better substrate for RMP than for chymosin. Milk-clotting activity towards κ-casein for RMP (0.4 μg^{-1}) is only 1/6 of that of chymosin (2.5 μg^{-1}). Therefore, the milk-clotting activity for RMP is still low compared to that of chymosin.

1.3.3 The inhibition of RMP by pepstatin A

The inhibition of RMP by pepstatin A was first reported by Rickert and McBride-Warren (1977). RMP is progressively inhibited by increasing the amounts of pepstatin A when incubated at pH 5.0 for 30 minutes at 25°C.

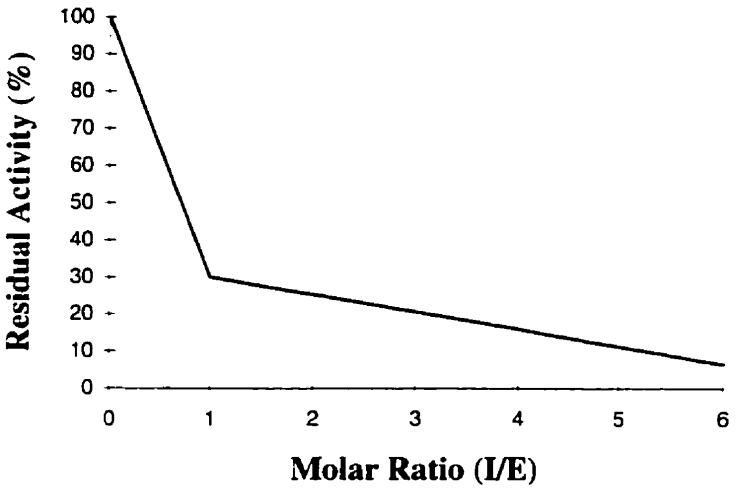


Figure 1.4. The progressive inhibition of RMP by pepstatin A. This diagram is a redrawing of the original diagram by Rickert & McBride-Warren (1977)

The inhibition curve is shown in Figure 1.4. From the inhibition diagram, it is obvious that the inhibition of RMP by pepstatin A is not in a 1:1 ratio. RMP quickly loses 70% of its activity as the molar ratio of pepstatin A : RMP increases from 0 to about 1.0. At this rate, the enzyme would be expected to be 100% inhibited at the pepstatin A : RMP molar ratio of 1.5. However, the inhibition reaction undergoes a dramatic change after RMP loses 70% activity. The residual 30% activity seems to be resistant to pepstatin A's inhibition. Therefore, 15 times more pepstatin A is required to eliminate the residual 30%

activity of RMP, and 100% inhibition is achieved at the pepstatin A : RMP molar ratio of 6. This type of inhibition by pepstatin A has also been observed in the studies of yeast protease A (Saheki & Holzer, 1975) and kidney renin (Miller *et al.*, 1972).

Rickert and McBride-Warren (1977) also carried out the studies on the conformational change associated with pepstatin A's inhibition by monitoring with CD spectroscopy. They observed that there was no major conformational change of RMP when the enzyme was incubated with pepstatin A at an equimolar amount. The spectrum suggested that there was only a minor general "loosening" of the tertiary structure of RMP. When RMP was incubated with sufficient amount of pepstatin A to totally inhibit the enzyme, a major conformational change was observed. The CD spectrum suggested a relaxation of the distorted β-structure predicted for the native RMP structure (McBride-Warren & Rickert, 1973).

Schechter and Berger (1967) pointed out that the substrate-binding site of proteinases could be divided into several subsites, each accommodating one amino acid residue of the peptide substrates. These subsites are located on both sides of the catalytic site and are named as S1, S2, S3, *etc.* and S1', S2', S3', *etc.* (Figure 1.5). The corresponding residues for each of the binding

subsites in the peptide substrates/inhibitors are named as P1, P2, P3, *etc.* and P1', P2', P3', *etc.* For the inhibitor pepstatin A molecule, the positional labels for the residues are shown in Figure 1.6. Because statine is a dipeptide analogue, it occupies two positions. The binding subsites in the RMP molecule are from S4 to S4'.

1.3.4 Other characteristics of RMP

RMP is the most highly glycosylated among the aspartic proteinases, possessing about 6% carbohydrate (Rickert & McBride-Warren, 1974). The carbohydrates are N-acetyl-D-glucosamine, D-mannose, D-glucose and D-galactose at the ratio of 2:7:2:1 (Rickert & McBride-Warren, 1974). The glycosylation sites of RMP are Asn79 and Asn188, respectively (Boel *et al.*, 1986). Both of the carbohydrate-enzyme linkages are type I linkages, i.e., N-acetyl-D-glucosamine linked to an asparagine residue within the sequence Asn-X-Thr (X is any of the common amino acids) which is called a sequeon (Devlin, 1982). Oxidation of RMP with 0.1 M periodate for 5 hours at 0°C results in a 40% loss of carbohydrate and no activity loss (McBride-Warren & Rickert, 1973). Prolonged oxidation of RMP (up to 48 hours) gives additional 10% loss in carbohydrate content and 50% decrease of enzyme activity. The residual carbohydrate is resistant to further oxidation. A thermal denaturation

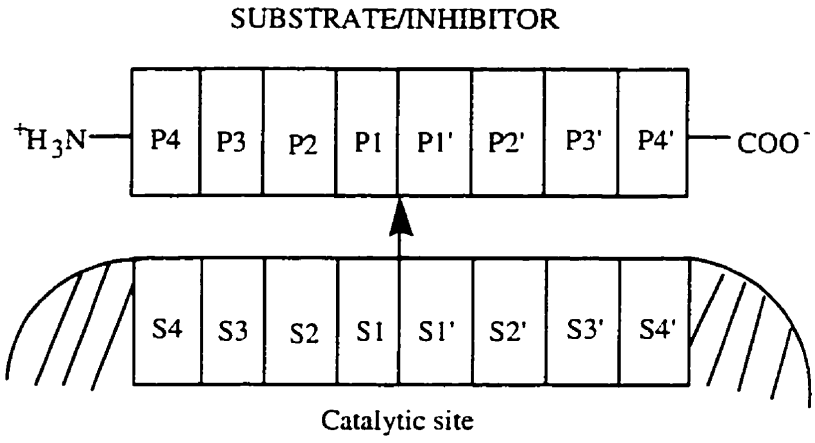


Figure 1.5. Schematic representation of the substrate binding site in the aspartic proteinases.

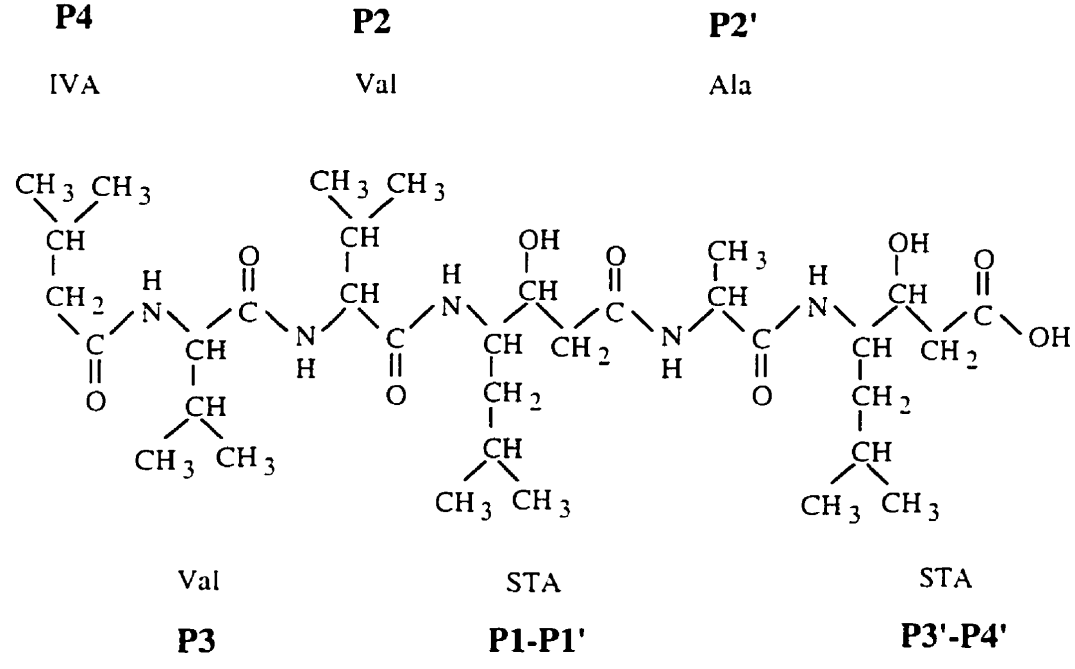


Figure 1.6. The positional names for the residues in the pepstatin A molecule.

study (Ledward *et al.*, 1975) indicated that RMP has maximum stability (77°C) at its isoelectric point. Studies by Hyslop *et al.* (1979) indicates that RMP is the most thermostable aspartic proteinase. This high thermal stability of RMP limits its use as a milk coagulant. In 1991, Brown and Yada reported that the native RMP was more thermally stable than the partially deglycosylated RMP (37.8% less carbohydrate) in heat denaturation studies. This result suggests that N-linked carbohydrates on RMP help to provide the enzyme with high thermal stability.

CD spectroscopy shows that RMP consists mainly of β-structures and few helical segments (McBride-Warren & Rickert, 1973). CD spectra of the oxidized RMP indicates that there are no major conformational changes up to 5 hours after the exposure to periodate. Decrease of β-structure is observed for the prolonged oxidized RMP (24-48 hours). Studies by De Konning & Draaisma (1973) and Lagrange *et al.* (1980) show that the isoelectric point for RMP is 4.58. RMP can be rapidly inhibited at pH 5.0 and 10°C by a 78-fold molar excess of diazoacetyl norleucine methyl ester while adding a 78-fold molar excess of Cu^{2+} simultaneously (Rickert & McBride-Warren, 1977). This indicates that RMP can also inhibited by diazonium compounds. Denaturation of RMP by Gdn-HCl (Brown & Yada, 1991) is described as a two-stage model

where only the native and denatured molecules exist in significant concentrations.

1.4 Aims and objectives

1.4.1 The structural determination of RMP

Our understanding of biological function of a protein molecule relies on our knowledge of its three-dimensional structure. From the three-dimensional structure, precise knowledge of the protein fold, and the conformations of the active site, substrate-binding site, regulatory site *etc.* can be obtained, and the reaction mechanism can be proposed if the protein is an enzyme. From the structure and proposed reaction mechanism, modifications of the protein molecule can be carried out to try to improve the molecule's biological activities and remove unwanted side effects. X-ray crystallography is one of the major methods in determining three-dimensional structures of proteins.

Modifications of RMP have been undertaken to try to improve its milk-clotting activity and decrease its thermal stability. However, these modifications are still in the trial and error stage without the guidance of a three dimensional structure. With RMP's increasing share of the world milk-clotting enzyme market (about 30% now; Harboe, 1996), structure-based modifications of this enzyme have become a priority. Therefore, the crystal structural determination of RMP was carried out. From the crystal structure,

the active site conformation, substrate-binding type and possible catalytic mechanism of RMP can be obtained. The preliminary studies of RMP were carried out by Jia *et al.* (1995). Jia *et al.* (1995) established the initial crystallization conditions of RMP and obtained a preliminary molecular replacement solution of RMP with an incomplete 2.8 Å data set of poor quality, using the coordinates of *Mucor pusillus* aspartic proteinase (MPP) as a search model. In the research project, crystals with better qualities were grown under improved crystallization conditions and a good quality 2.15 Å synchrotron data set was collected at DESY synchrotron, EMBL, Hamburg. From this crystallographic study of RMP, the following questions were also expected to be answered in order to provide a guideline for further modification of RMP.

(1) Why is RMP effective as a milk-clotting enzyme?

(2) Why does RMP have the greatest thermal stability of any milk-clotting aspartic proteinase?

(3) What is the function of the carbohydrate?

(4) How could RMP be modified to improve its properties?

(5) What determines the pH optimum for the activity of an aspartic proteinase?

1.4.2 The structural determination of the RMP-pepstatin A complex

After the crystal structure of RMP was solved, a structural comparison study of RMP with other aspartic proteinases was carried out (Yang *et al.*, 1997). The comparison results show that RMP and MPP diverged from the main stream of aspartic proteinases at an early stage of evolution. They form a sub-family of the aspartic proteinases. Since no structures of RMP or MPP complexes with inhibitors were published at the time that the native RMP structure was solved, a study of the RMP-pepstatin A complex was carried out. From the study of the RMP-pepstatin A complex, it was hoped to determine whether RMP has any differences in the substrate binding, catalytic mechanism or protein folding compared to other aspartic proteinases. This would improve our knowledge about aspartic proteinases. The results from the study of this complex may also provide a guideline for the synthesis of new potent aspartic proteinase inhibitors.

Chapter 2. Introduction to protein crystallography

The biological functions of a protein molecule are closely related to its three-dimensional structure. The three-dimensional structure of protein molecules can be determined by several methods, such as protein crystallography, macromolecular NMR and electron microscopy. Protein crystallography is the application of the techniques of X-ray diffraction to protein crystals. During the past thirty years, protein crystallography has advanced rapidly with the development of computer science. It has become one of the most powerful methods in elucidating the three-dimensional structure of proteins. The X-ray analysis of a protein structure usually includes the following four steps: crystallization, crystal diffraction and data collection, phase determination, and model-building and refinement. This chapter is a brief introduction to each of these steps. Most of the information provided in this chapter comes from the X-ray protein crystallography textbook: Principles of Protein X-ray Crystallography (Drenth, 1994). The following textbooks are also used as references: Protein Crystallography (Blundell & Johnson, 1976), Crystal Structure Analysis: A Primer (Glusker & Trueblood, 1985) and X-ray Structure Determination: A Practical Guide (Stout

& Jensen, 1989). As the key step of protein crystallography is to determine the phase angle for each reflection, and the molecular replacement method (MR) was used to determine the phase angles for RMP, this method will be discussed in greater detail.

2.1 Crystallization of proteins

Crystals are formed by the three-dimensional periodic repeat of a building block. This building block is called a unit cell. The unit cell is characterized by six parameters: three axial lengths: *a*, *b*, and *c*; and three interaxial angles α, β, and γ. According to these six parameters, the crystals are divided into seven crystal systems (Table 2.1). A unit cell contains one or more molecules related by crystallographic symmetries which can be defined by space groups. A space group is a group of symmetry operations consistent with an infinitely extended, regularly repeating pattern. There are a total of 230 space groups. Because they are chiral molecules, proteins can crystallize only in space groups that have only rotational and/or translational symmetry elements. The most common space group found for proteins is $P2_12_12_1$.

Compared to X-ray protein crystallography, protein crystallization is an underdeveloped area and is still at a trial-and-error stage. Crystallization databases, screens and robots have been developed. However, the determination of optimum crystallization conditions for proteins is still a matter

Table 2.1. The seven crystal systems.

Crystal system	Number of independent parameters	Parameters	Reciprocal lattice symmetry
Triclinic	6	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$	$\bar{1}$
Monoclinic	4	$a \neq b \neq c$; $\alpha = \gamma = 90°$; $\beta > 90°$	2/m
Orthorhombic	3	$a \neq b \neq c$; $\alpha = \beta = \gamma = 90°$;	mmm
Tetragonal	2	$a = b \neq c$; $\alpha = \beta = \gamma = 90°$	4/mmm
Trigonal			
rhombohedral	2	$a = b = c$; $\alpha = \beta = \gamma \neq 90°$	$\bar{3}$m
hexagonal	2	$a = b \neq c$; $\alpha = \beta = 90°$; $\gamma = 120°$	6/mmm
Hexagonal	2	$a = b \neq c$; $\alpha = \beta = 90°$; $\gamma = 120°$	6/mmm
Cubic	1	$a = b = c$; $\alpha = \beta = \gamma = 90°$	m3m

of trial and error. The crystallization of proteins usually involves four important steps (Drenth, 1994):

(1) Determination of the purity of the protein. Usually the purer the protein, the better the chances to grow crystals.

(2) The protein is dissolved in a suitable solvent from which the protein crystal is grown. Usually the solvent is a water-buffer solution,

sometimes with an organic solvent such as 2-methyl-2,4-pentanediol (MPD) added. Normally, the precipitant is also added to an extent that it does not cause precipitation of the protein.

(3) The protein solution is equilibrated with the precipitant solution to reach a supersaturation stage. Spontaneous formation of nuclei is achieved at this high supersaturation state.

(4) Once the nuclei are formed, protein crystals start to grow.

To achieve crystal growth, supersaturation should be kept at a lower level so that only a few nuclei are formed. Crystals should grow slowly to reach a maximum degree of order in their structure. Although many techniques have been developed for protein crystallization, they fall into one of the following four major categories:

(1) Batch crystallization. The precipitating reagent is added to the protein solution to raise the protein solution to a high supersaturation state, from which the crystals may grow.

(2) Liquid-liquid diffusion. In this method the protein solution and the solution containing the precipitant are layered on top of each other in a small-pore capillary tube. The solution with a higher density is at the bottom.

(3) Vapor diffusion method. This is the most common method. In this method, the protein solution is equilibrated with the precipitating solution through the vapor phase. The protein concentration is slowly increased to a supersaturation state. This technique can be divided into three methods: hanging drop method, sitting drop method and sandwich-drop method.

(4) Dialysis. In this method, the protein solution is dialyzed against the precipitating solution to achieve nuclei formation and crystal growth. The advantage of this method is that one can change the precipitating solution readily.

2.2 Data collection and data processing

X-rays were discovered by German physicist Wilhelm C. Roentgen in 1895. The diffraction of X-rays by crystals was discovered by Max von Laue in 1912. He described the diffraction from a crystal as diffraction from a three-dimensional grating. In the same year, W. L. Bragg, the father of modern crystallography, noticed the similarity of diffraction to ordinary reflection (Figure 2.1) and deduced a simple equation (Bragg's law) treating the diffraction as "reflection" from planes in the lattice.

$$n\lambda = 2d_{hkl}\sin\theta_{hkl}$$

where n is an integer, λ is the wavelength of the X-ray radiation used, d_{hkl} is the perpendicular distance between adjacent lattice planes determined by the Miller indices *hkl*, and θ_{hkl} is one half of the angle between the incident and diffracted X-ray beams for reflections from these planes.

The intensities *I*(*hkl*)s of the reflections can be measured by several ways: single photon scintillation counters, photographic films, image plates and area detectors. For protein crystallography, the intensities of reflections

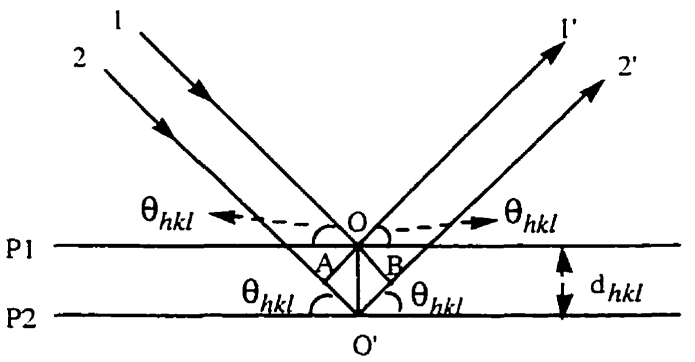


Figure 2.1. Construction showing Bragg's law for diffraction. P_1 and P_2 are a pair of reflection planes with interplanar spacing d_{hkl}. 1 and 2 are the parallel incident beams which make an angle θ_{hkl} with the reflection planes.1' and 2' are the diffracted beams. O'A + O'B is the path length difference between beams 1' and 2', which is equal to $2d_{hkl}\sin\theta_{hkl}$.

are usually measured by the last two methods. Image plates are plates covered with a thin layer of inorganic storage phosphor. When exposed to X-rays, the X-ray photons excite electrons in the phosphor to higher energy states. Part of this energy is emitted very quickly as fluorescence light in the visible wavelength region. However, a significant amount of energy is stored in the colour centers of the phosphor for a period of several days, and is slowly released. On illumination with strong white light, the energy trapped in the colour centers is released very quickly. In practical applications, the image plates are scanned with a red laser and the emitted blue light is

measured with a photomultiplier. The light emitted is usually proportional to the number of photons to which that particular region of the plate was exposed. Area detectors are electronic devices that detect X-ray photons on a two-dimensional surface and process the signal immediately after photon detection. At present, there are three types of area detectors: gas-filled ionization type detectors, video-based area detectors and charge coupled device (CCD) type detectors.

The diffraction from a crystal can be described as a summation of the enormous number of waves scattered by the electrons in the crystal. Because the electrons are bound to atoms, each type of atom is assigned an atomic scattering factor f. The atomic scattering factor f is a function of $(\sin\theta/\lambda)$, where θ is the diffraction angle and λ is the X-ray wavelength. f is equal to the number of electrons in the atom at $\theta = 0$, and it decreases with increasing θ angle. Next let us consider the case of scattering by a unit cell (Figure 2.2). Suppose a unit cell has n atoms at position $\mathbf{r}_j$ ($j = 0, 1, 2, \ldots, n$) with respect to the origin of the unit cell, and each atom has atomic scattering factor as f_j. The total scattering by the unit cell ($\mathbf{F}(\mathbf{S})$) is

$$\mathbf{F}(\mathbf{S}) = \sum_{j=1}^{n} f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]$$

$\mathbf{F}(\mathbf{S})$ is the structure factor. It is dependent on the arrangement of the atoms in the unit cell. $\mathbf{S}$ is the vector perpendicular to the imaginary "reflecting plane" and makes equal angles with the incident and reflected beams.

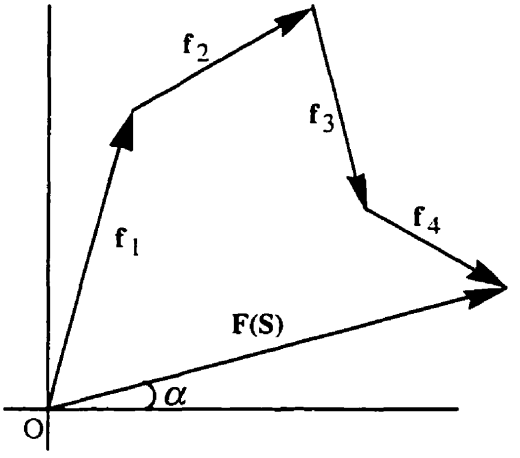


Figure 2.2. The structure factor $\mathbf{F}(\mathbf{S})$ is the summation of the scattering of all the atoms in the unit cell. $\mathbf{f}_j$s are the scattering vectors of the atoms in the Argand diagram with respect to the origin of the unit cell ($\mathbf{f}_j = f_j \exp[2\pi \mathbf{r}_j \cdot \mathbf{S}]$).

From the diffraction experiment, only $I(hkl)$s are obtained. $I(hkl)$s have to be converted to the structure factor amplitudes $|F(hkl)|$s because it is $|F(hkl)|$s that are used in the Fourier transformation to get the electron density map.

The relationship between $I(hkl)$ and $|F(hkl)|$ is:

$$I(hkl) = \frac{\lambda^3}{\omega \cdot V^2} \cdot \left(\frac{e^2}{mc^2}\right)^2 \cdot V_{cr} \cdot I_0 \cdot L \cdot P \times A \cdot |F(hkl)|^2$$

where λ is the wavelength, ω is the angular velocity that the crystal is rotated, V is the volume of the unit cell, V_{cr} is the volume of the crystal, L is the Lorentz factor which depends on the data acquisition technique, P is the polarization factor, A is the absorption factor, e and m are the charge and mass of an electron, c is the velocity of light, and I_0 is the intensity of the incident beam.

If we assign K equal to:

$$\frac{\omega \cdot V^2}{\lambda^3} \cdot \left(\frac{mc^2}{e^2}\right)^2 \cdot \frac{1}{V_{cr} \times I_0}$$

then,

$$|F(hkl)| = \left(\frac{KI(hkl)}{L \cdot P \cdot A}\right)^{1/2}$$

Because the scale factor K is normally a constant for any given set of measurements, it is commonly omitted for the data reduction calculations. Therefore, the structure factor amplitude $|F(hkl)|$ (also called the observed structure factor amplitude $|F_o|$) is obtained on a relative scale by calculating

$I(hkl)/(L \cdot P \cdot A)$. The Lorentz and polarization corrections are usually incorporated into the software package for processing the intensity data. Whether correction for absorption is required or not, depends on the shape of the crystal, the wavelength of the X-rays, and the diffraction technique.

2.3 Phase determination

The electron density in a crystal can be obtained by the Fourier transformation

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}|F(hkl)|\exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

in which $|F(hkl)|$ is the structure factor amplitude of reflection (hkl), including the temperature factor, and $\alpha(hkl)$ is the phase angle. x, y, and z are the coordinates in the unit cell. From the previous section, we know that the amplitudes $|F(hkl)|$ can be obtained from the $I(hkl)$ after applying the correction factors of L, P and A. However, the phase angles $\alpha(hkl)$ are not available. In general, the phases for the reflections can be obtained by one of the following methods in protein X-ray crystallography:

(1) Multiple isomorphous replacement method (MIR). This method requires the attachment of heavy atoms to the protein molecule in the crystal.

(2) Molecular replacement method (MR). This method is based on the similarity of the unknown structure to an already known structure.

(3) Multiple wavelength anomalous diffraction method (MAD). In this method, strong anomalously scattering atoms should be present in the protein structure.

(4) Direct methods. The application of direct methods to protein crystal structural determination is still at the preliminary stage.

Because the structure of RMP was solved by the molecular replacement method, only that method will be discussed in this chapter.

2.4 Molecular replacement (MR)

Molecular replacement is the most rapid method for solving protein structures. In this section, two textbooks, Protein Crystallography (Blundell & Johnson, 1976) and Principles of Protein X-ray Crystallography (Drenth, 1994), are used as references to give a brief introduction to the molecular replacement method. This technique requires that the protein with unknown structure and a protein whose structure is known be homologous. The known structure of this homologous protein is used as a model for the protein for which the structure is to be determined. This method is based on the observations that proteins, which have a high homology in amino acid sequence, often have homologous three-dimensional structures.

The molecular replacement method was initiated in pioneering studies by Rossmann and Blow (1962). This method transfers the known molecular structure of a protein in its native crystal structure to the crystal of the protein for which the structure is unknown. This task is carried out in two steps: rotation and translation. The rotation is to put the unknown and known proteins into the same spatial orientation and the translation is used to

superimpose the correctly-oriented molecules. The molecular replacement method can also be used for noncrystallographic symmetry.

In both rotation and translation functions, the Patterson functions of the protein molecules are used. The Patterson function $P(\mathbf{u})$ is a Fourier summation with amplitudes $|F(\mathbf{S})|^2$ as coefficients and all phase angles equal to zero.

$$P(\mathbf{u}) = \frac{1}{V}\sum_{S}|F(\mathbf{S})|^2 \exp(-2\pi i\mathbf{u}\cdot\mathbf{S})$$

where V is the volume of the unit cell.

The Patterson map has peaks at end points of vector $\mathbf{u}$ equal to vectors between atoms in the unit cell. Therefore, the Patterson map is a vector map and is centrosymmetric. If the pair of atoms belong to the same molecule, the vector between these two atoms is called a self-Patterson vector. If the pair of atoms belong to two different molecules, the vector is called a cross-Patterson vector. The self-Patterson vectors are used in the rotation search and the cross-Patterson vectors are used in the translation search in the molecular replacement method.

2.4.1 The rotation function

The self-Patterson peaks of a molecule all lie in a volume around the origin with a radius equal to the dimension of the molecule. If non-crystallographic symmetry is present in the crystal, the self-Patterson vectors for each of the molecules in the asymmetric unit will be the same except for a rotation that is the same as their noncrystallographic rotational symmetry. Therefore, maximum overlap between two self-Patterson maps can be obtained if the two Patterson functions are superimposed on a correct rotation. Similarly, the correct rotation of the self-Patterson function of one molecule onto the self-Patterson function of another molecule in a different crystal lattice will result in a maximum overlap between the two Patterson maps.

Let us consider one atom located at position $\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$ in one crystallographic lattice with axes $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$. Rotation of the lattice around its origin leads to a new set of axes: $\mathbf{a}_{r.1}$, $\mathbf{a}_{r.2}$, and $\mathbf{a}_{r.3}$. The position of the atom in this new system is $\mathbf{x}_r = x_{r.1}\mathbf{a}_{r.1} + x_{r.2}\mathbf{a}_{r.2} + x_{r.3}\mathbf{a}_{r.3}$ and the relationship between the two sets of coordinates is $\mathbf{x}_r = [C]\ \mathbf{x}$, where [C] is the rotation matrix. A rotation of the axes has the same effect as a rotation of the molecule in the opposite direction. The Patterson map rotates as the molecule rotates.

Applying the rotation matrix [C] to the Patterson function $P(\mathbf{u})$ gives the rotated Patterson function $P_r(\mathbf{u}_r)$. An overlap function R of $P(\mathbf{u})$ with the rotated $P_r(\mathbf{u}_r)$, of the same crystal lattice (self-rotation function) or a different crystal lattice (cross-rotation function) is defined as:

$$R(\alpha,\beta,\gamma) = \int_U P(\mathbf{u}) \times P_r(\mathbf{u}_r) d\mathbf{u}$$

U is the volume in the Patterson map where all the self-Patterson vectors are located. α, β and γ are the Eulerian rotation angles (Machin, 1985). The rotation function R depends on the rotation angles (related to [C]) and will reach a maximum value if the Patterson maps are correctly overlapped.

Now $P(\mathbf{u})$ and $P_r(\mathbf{u}_r)$ are expanded as a Fourier series:

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp[-2\pi i \mathbf{h}\mathbf{u}]$$

and

$$P_r(\mathbf{u}_r) = \frac{1}{V}\sum_{\mathbf{h}'} |F(\mathbf{h}')|^2 \exp[-2\pi i \mathbf{h}'\mathbf{u}_r]$$

Because $\mathbf{u}_r = [C]\,\mathbf{u}$,

$$P_r(\mathbf{u}_r) = \frac{1}{V}\sum_{\mathbf{h}'} |F(\mathbf{h}')|^2 \exp[-2\pi i \mathbf{h}'[C]\mathbf{u}]$$

As $\mathbf{h}'[C]$ equals to $[C^{-1}]\mathbf{h}'$, therefore we get

$$P_r(\mathbf{u}_r) = \frac{1}{V}\sum_{\mathbf{h}'}|F(\mathbf{h}')|^2 \exp[-2\pi i[C^{-1}]\mathbf{h}'\mathbf{u}]$$

Now the rotation function $R(\alpha,\beta,\gamma)$ is:

$$R(\alpha,\beta,\gamma) = \frac{1}{V^2}\sum_{\mathbf{h}}\sum_{\mathbf{h}'}|F(\mathbf{h})|^2|F(\mathbf{h}')|^2 \cdot \int_U \exp[-2\pi i(\mathbf{h} + [C^{-1}]\mathbf{h}')\mathbf{u}]d\mathbf{u}$$

The integral is an interference function and its value may be written as:

$$(U/V)\ \mathbf{G}[-(\mathbf{h} + [C^{-1}]\ \mathbf{h}')]$$

The integration is over a volume symmetrical about the Patterson origin. Therefore, the rotation function is:

$$R(\alpha,\beta,\gamma) = \frac{U}{V^3}\sum_{\mathbf{h}}\sum_{\mathbf{h}'}|F(\mathbf{h})|^2|F(\mathbf{h}')|^2 \times \mathbf{G}[-(\mathbf{h} + [C^{-1}]\mathbf{h}')]$$

If the protein molecule can be assumed to be approximately spherical, the function G can be written as:

$$G = \frac{3\,(\sin 2\pi x - 2\pi x \cos 2\pi x)}{(2\pi x)^3}$$

x is equal to $(\mathbf{h} + [C^{-1}]\mathbf{h}') \cdot \mathbf{r}$.

The function G is shown in Figure 2.3. It has a maximum value of 1 and is less than 0.086 outside the range $-0.725 < x < 0.725$. Therefore, all the terms in the rotation function for which $|x| > 0.725$ are neglected.

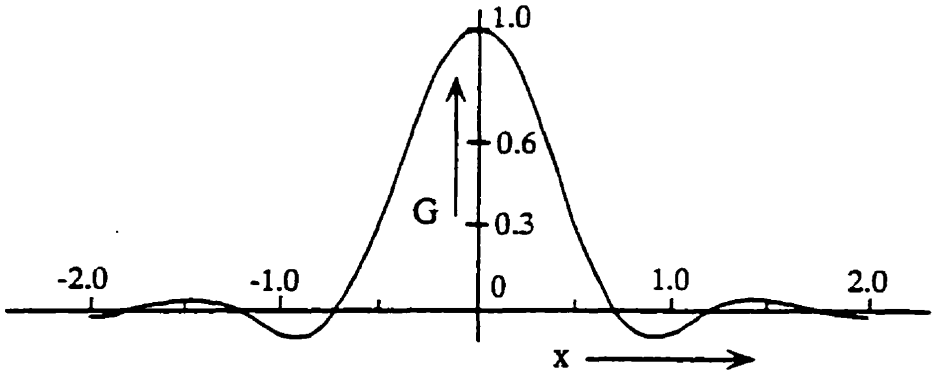


Figure 2.3. The function $G = [3(\sin 2\pi x - 2\pi x \cos 2\pi x)]/(2\pi x)^3$ plotted as a function of x (Rossmann & Blow, 1962).

Because the rotation function is approximately proportional to $|F|^4$, all strong reflections should be present in the calculation. Very low resolution data (> 10 Å^2) can be excluded in the calculation because they are determined mostly by the solvent region and insensitive to rotation.

2.4.2 The translation function

From the rotation function, the orientation of the unknown molecule is determined. No translation is incorporated. However, for the final result of the molecular replacement method, the position of the unknown molecule in the

unit cell has to be determined. This is done by a translation which overlaps one molecule onto the other in real space. One way to do this is by trial and error. The known molecule is moved through the asymmetric unit and structure factors (F_c) are calculated and compared with the observed structure factors (F_o) by calculating an R-factor or the correlation coefficient as a function of the molecular position.

$$R = \frac{\sum_{hkl} \| F_o | - k | F_c \|}{\sum_{hkl} | F_o |}$$

or

$$C = \frac{\sum_{hkl} \left(|F_o|^2 - \overline{|F_o|^2} \right) \cdot \left(|F_c|^2 - \overline{|F_c|^2} \right)}{\left[\sum_{hkl} \left(|F_o|^2 - \overline{|F_o|^2} \right)^2 \sum \left(|F_c|^2 - \overline{|F_c|^2} \right)^2 \right]^{1/2}}$$

This correlation coefficient is a better index than the R-factor because it is scaling insensitive; the replacement of $|F_o|^2$ by $k|F_o|^2$ + a constant (k is the scale factor for the intensities) gives the same value for the correlation coefficient.

In a more straightforward method than the trial and error search, a translation function is investigated. In this method, the correlation between a set of cross-Patterson vectors for a model structure and the observed Patterson function is calculated. The cross-Patterson vectors are the intermolecular vectors.

Let us assume that crystals of the unknown structure belong to a space group which has at least two equivalent positions. Consider a pair of molecules which are symmetry-related (Figure 2.4). Their orientations are known from the rotation function, but not their positions in the unit cell. From the translation function, the position of molecule 1 with respect to symmetry-related molecule 2 can be determined, and subsequently the relative position of any other pair of symmetry-related molecules can be determined. In space group P1, no crystallographic symmetry exists. The origin can be chosen anywhere and this has no influence on the absolute value of the structure factor. Therefore, the calculation of the translation function is not necessary for triclinic protein crystals.

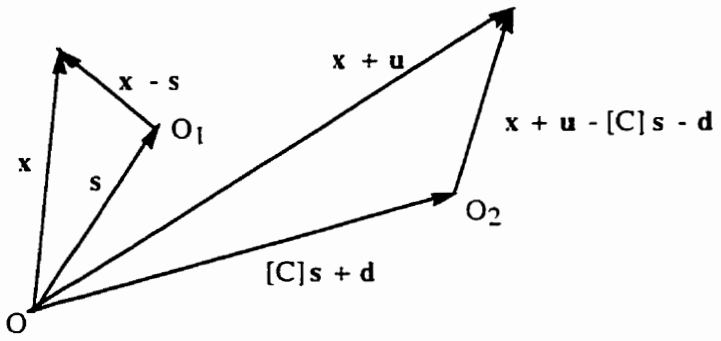


Figure 2.4. The position (**x**) in molecule 1 and (**x** + **u**) in molecule 2. O_1 and O_2 are the local origins.

The origin of the unit cell is at O. The local origins of molecules 1 and 2 are O_1 and O_2, respectively. The cross-Patterson function between these two molecules is:

$$P_{1,2}(\mathbf{u}) = \int_V \rho_1(\mathbf{x}) \times \rho_2(\mathbf{x}+\mathbf{u}) d\mathbf{x}$$

If the electron density expressed with respect to the local origin of the first molecule, the model molecule M, is ρ_M, then

$$\rho_1(\mathbf{x}) = \rho_M(\mathbf{x}-\mathbf{s})$$

and

$$\rho_2(\mathbf{x}+\mathbf{u}) = \rho(\mathbf{x}+\mathbf{u}-[C]\mathbf{s}-\mathbf{d})$$

for the local origin in O_2. This is equal to the electron density in the model molecule (molecule 1) at the symmetry-related position $[C^{-1}](\mathbf{x}+\mathbf{u}-[C]\mathbf{s}-\mathbf{d})$. [C] is the matrix that transfers molecule 1 to molecule 2, and $[C^{-1}]$ is the inverse matrix of [C]. Therefore,

$$\rho_2(\mathbf{x}+\mathbf{u}) = \rho_M\left\{[C^{-1}](\mathbf{x}+\mathbf{u}-[C]\mathbf{s}-\mathbf{d})\right\}$$

Now the cross-Patterson function is:

$$P_{1,2}(\mathbf{u},\mathbf{s}) = \int_V \rho_M(\mathbf{x}-\mathbf{s}) \times \rho_M\left\{[C^{-1}](\mathbf{x}+\mathbf{u}-[C]\mathbf{s}-\mathbf{d})\right\} d\mathbf{x}$$

If the intermolecular vector **t** is between O_1 and O_2,

$$\mathbf{t} = -\mathbf{s} + [C]\mathbf{s} + \mathbf{d}$$

Now we write ρ_M as a Fourier series in terms of the structure factor $\mathbf{F}_M$ and then simplify the cross-Patterson function. The simplified cross-Patterson function is:

$$P_{1,2}(\mathbf{u},\mathbf{t}) = \sum_{\mathbf{h}} \mathbf{F}_M(\mathbf{h}) \cdot \mathbf{F}_M^*(\mathbf{h}[C]) \exp[-2\pi i \mathbf{h} \cdot \mathbf{t}] \exp[-2\pi i \mathbf{h} \cdot \mathbf{u}]$$

This is the cross-Patterson function of the model structure in which two molecules are related by crystallographic symmetry. $\mathbf{F}_M^*$ is the complex conjugate of $\mathbf{F}_M$.

The translation function $T(\mathbf{t})$ is calculated as:

$$T(\mathbf{t}) = \int_V P_{1,2}(\mathbf{u},\mathbf{t}) \times P(\mathbf{u}) d\mathbf{u}$$

The translation function $T(\mathbf{t})$ will reach a maximum if the intermolecular vector $\mathbf{t}$ is equal to the true intermolecular vector $\mathbf{t}_0$. If we put the Fourier expansion form of $P_{1,2}(\mathbf{u},\mathbf{t})$ into the translation function and replace $P(\mathbf{u})$ with $\sum_{\mathbf{p}} |F_o(\mathbf{p})|^2 \exp[-2\pi i \mathbf{p}\cdot\mathbf{u}] d\mathbf{u}$, we get the translation function as:

$$T(\mathbf{t}) = \int_V \sum_{\mathbf{h}} \mathbf{F}_M(\mathbf{h}) \cdot \mathbf{F}_M^*(\mathbf{h}[\mathrm{C}]) \exp[-2\pi i \mathbf{h}\cdot\mathbf{t}] \exp[-2\pi i \mathbf{h}\cdot\mathbf{u}] \times \sum_{p} |F_o(\mathbf{p})|^2 \exp[-2\pi i \mathbf{p}\cdot\mathbf{u}] d\mathbf{u}$$

$$= \sum_{\mathbf{h}} F_M(\mathbf{h}) \cdot \mathbf{F}_M^*(\mathbf{h}[\mathrm{C}]) \exp[-2\pi i \mathbf{h}\cdot\mathbf{t}] \times \sum_{p} |F_o(\mathbf{p})|^2 \int_V \exp[2\pi i (\mathbf{h}-\mathbf{p})\cdot\mathbf{u}] d\mathbf{u}$$

The integral vanishes unless $\mathbf{h}$-$\mathbf{p}$=0 or $\mathbf{h}$=$\mathbf{p}$. Therefore, the translation function is re-written as:

$$T(\mathbf{t}) = \sum_{\mathbf{h}} |F_o(\mathbf{h})|^2 \cdot \mathbf{F}_M(\mathbf{h}) \cdot \mathbf{F}_M^*(\mathbf{h}\cdot[\mathrm{C}]) \exp[-2\pi i \mathbf{h}\cdot\mathbf{t}]$$

This Fourier summation is the final form of the translation function. Because all the coefficients in the summation are known, the translation function can be easily calculated.

2.5 Refinement of the model

When an approximate initial model is obtained by MIR, MAD or MR, the model is subjected to refinement. A crystallographic R-factor is usually used as an index of closeness between the model and the real structure. The R-factor is calculated by the following equation:

$$R = \frac{\sum_{hkl} \| F_o | - k | F_c \|}{\sum_{hkl} | F_o |}$$

in which, $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes for reflection (hkl), respectively. k is a scale factor.

In the refinement, the model is adjusted so that a closer agreement between the calculated and observed structure factors can be reached. Although many computer programs are available for structure refinement, they all fall into one of the two common refinement methods: least-squares refinement and simulated-annealing refinement. All these refinement programs use empirical restraints or constraints to ensure that a reasonable structure ensues during the refinement steps. In the refinement, the positional

parameters and the temperature factors for all the atoms except hydrogen atoms are adjusted. Usually only the hydrogen atoms linked to polar atoms are placed by geometry on the corresponding non-hydrogen atoms and assigned fixed temperature factors (e.g. 15 $Å^2$). Because Powell energy minimization with the conjugate gradient method (Powell, 1977) was used in the refinement of the structures of RMP and its pepstatin A complex, a brief introduction to this method is given in this section. For a well-refined protein structure, the *R*-factor is usually less than 25%. However, it has been found that some incorrectly-built protein structural models can have similar *R*-factors. To detect this situation, Brünger (1992a, 1993) introduced a cross-validation scheme based on a new index called *R*-free. In this method, the unique reflections are divided into two groups: a "test set (T)" and a "working set (W)". The test set is a random selection of about 5 ~ 10% of the observed reflections. The refinement is carried out with the working set only, and the *R*-free is calculated with only the test set of reflections by the following equation:

$$R\text{-free} = \frac{\sum_{hkl \subset T} \| F_o | - k | F_c \|}{\sum_{hkl \subset T} | F_o |}$$

where $hkl \subset T$ means all the reflections in the test set T. Thus, *R*-free is unbiased by the refinement process.

2.5.1 Least-squares refinement

The least-squares refinement in protein crystallography is based on the principle of least-squares. From the data reduction step, the amplitude of structure factor $|F_o(hkl)|$ can be obtained from $I(hkl)$. From the preliminary model, values for the structure factors $\mathbf{F}_c(hkl)$ can be calculated. The least-squares refinement is used to bring the values of $|F_c(hkl)|$ as close as possible to $|F_o(hkl)|$ for all reflections (hkl) by varying the parameters of the model. The parameters that can be varied in the model are the atomic coordinates and the atomic or the overall model temperature factors. The function to be minimized in the least-squares refinement is the function Q.

$$Q = \sum_{hkl} w(hkl)\left(|F_o(hkl)| - |F_c(hkl)|\right)^2$$

The summation is over all crystallographically-independent reflections and $w(hkl)$ is the weight given to an observation. For small molecule crystallography, $w(hkl)$ is usually equal to $1/\sigma^2(hkl)$. For protein crystallography, $w(hkl)$ is a $sin\theta/\lambda$-dependent function. In the process of minimizing the function Q, the atomic parameters u_j (j = 1,2,3...n) that determine $|F_c(hkl)|$ are varied. The minimization of the function Q is carried out by setting the differentials of Q with respect to all u_j equal to zero: $\partial Q / \partial u_j = 0$ or

$$\sum_{hkl} w(hkl)(|F_o(hkl)| - |F_c(hkl)|)\frac{\partial |F_c(hkl)|}{\partial u_j} = 0$$

The observed structure factor amplitudes ($|F_o(hkl)|$) are constants and the amplitudes of the calculated structure factors ($|F_c(hkl)|$) are dependent on the variables u_j. To solve these equations, each $|F_c(hkl)|$ is expressed in a Taylor expansion:

$$|F_c(hkl;u)| = |F_c(hkl;u_s)| + \sum_i \varepsilon_i \left[\frac{\partial |F_c(hkl;u)|}{\partial u_i}\right]_{u_s}$$

$$+ \frac{1}{2}\sum_j \sum_i \varepsilon_j \varepsilon_i \frac{\partial^2 |F(hkl;u)|}{\partial u_j \partial u_i} + \cdots$$

$|F_c(hkl;u)|$ indicates that the $|F_c|$ depends on the parameters u. The initial values of u are u_s and are changed by a small amount ε. For parameter u_i,

$\varepsilon_i = u_i - u_{i,s}$. $\left[\partial|F_c(hkl;u)| / \partial u_i\right]_{u_s}$ is the differential of $|F_c(hkl;u)|$ with respect to u_i calculated at the starting value u_s. Because the ε values are very small, higher order terms in ε are neglected. Now putting the Taylor expansion form of $|F_c(hkl)|$ back into the differentials of Q equations, the so-called normal equations are obtained:

$$\sum_{hkl} w(hkl)(|F_o(hkl)| - |F_c(hkl;u_s|) \times \left[\frac{\partial|F_c(hkl;u)|}{\partial u_j}\right]_{u_s}$$

$$-\sum_i \varepsilon_i \sum_{hkl} w(hkl)\left[\frac{\partial|F_c(hkl;u)|}{\partial u_i}\right]_{u_s} \times \left[\frac{\partial|F_c(hkl;u)|}{\partial u_j}\right]_{u_s} = 0$$

With these n equations ($j = 1 \rightarrow n$), the ε values have to be found and applied to the variables u. With the truncation of higher order terms of the Taylor series the final results can be approached by iteration.

These normal equations are usually solved in the abbreviated forms: $\sum_i (\varepsilon_i a_{ij}) - b_j = 0$.

For $j = 1$: $a_{11}\varepsilon_1 + a_{21}\varepsilon_2 + a_{31}\varepsilon_3 + \cdots = b_1$

$j = 2$ $a_{12}\varepsilon_1 + a_{22}\varepsilon_2 + a_{32}\varepsilon_3 + \cdots = b_2$

These equations can be expressed in matrix form:

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & ---- \\ a_{12} & a_{22} & a_{32} & ---- \\ -- & -- & -- & ---- \\ -- & -- & -- & ---- \end{bmatrix} \times \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ - \\ - \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ - \\ - \end{bmatrix}$$

or [**A**] × [ε] = [**b**]. [**A**] is called normal matrix and its elements are

$$\sum_{hkl} w(hkl) \left[\frac{\partial |F_c(hkl;u)|}{\partial u_i} \right]_{u_r} \times \left[\frac{\partial |F_c(hkl;u)|}{\partial u_j} \right]_{u_r}$$

[ε] is the unknown vector. [**b**] is the gradient vector and its elements are:

$$\sum_{hkl} w(hkl)(|F_o(hkl)| - |F_c(hkl;u)|) \times \left[\frac{\partial |F_c(hkl;u)|}{\partial u_j} \right]_{u_r}$$

Above is the principle of the least-squares refinement used in X-ray crystallography.

2.5.2 Simulated Annealing

In the process of least-squares refinement, the refinement follows a downhill path to minimize the function. If the starting model is not very different from the real structure, the refinement can converge easily to the correct solution. However, if the starting model is quite far away from the real

structure, the refinement may be trapped in a local minimum instead of a global minimum state. To solve this problem, Brünger *et al.* (1987) introduced simulated-annealing refinement, which allows both downhill and uphill searching to try to ensure that the refinement can reach the real minimum. In this refinement, the dynamic behavior of a system of particles is simulated. This simulation yields an ensemble of structures that is energetically allowed for a given temperature and pressure. The energy distribution of the structures in the ensemble follows Boltzmann's law, i.e. the potential energy ε_{pot} is proportional to $\exp[-\varepsilon_{pot}/kT]$, where k is the Boltzmann constant and T is the absolute temperature. The potential energy depends on the relative positions of the atoms. In this method, the electrostatic term is also included. This term is defined as:

$$\sum \frac{q_i \times q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}$$

ε_0 is the permittivity constant; ε_r is the local dielectric constant, which is given an estimated value between 1 and 80; r_{ij} is the intermolecular distance between two atoms at i and j positions; q_i and q_j are the charges on these two atoms, respectively.

A molecular dynamics calculation on a molecule is initiated by assigning the atoms velocities which are derived from a Maxwell distribution at

an appropriate temperature. At time t=0, the atoms are in a starting configuration with a potential energy E_{pot} for the entire molecule. On each atom i at position $\mathbf{r}_i$, a force is acting. This force is derived from the potential energy: force (i) = $-\partial E_{pot} / \partial \mathbf{r}_i$. From the Newton equation it is known that the force (i) is equal to $m_i(d^2\mathbf{r}_i / dt^2)$ where m_i is the mass of atom i. Therefore, the acceleration $d^2\mathbf{r}_i / dt^2$ for atom i with mass m_i can be calculated and then applied. After a short time step Δt, in the femtosecond range (1 fsec. = 10^{-15} sec), the process is repeated with the atoms in the new positions. If the number of steps is sufficient (10^3-10^4), the minimum of E_{pot} is reached and information about the dynamic behavior of the atoms in the molecule is obtained. If the temperature of the system is raised to a higher value, more atoms will have a higher speed and a higher kinetic energy, and therefore can overcome higher energy barriers. The basic idea of this refinement method is to raise the temperature high enough so that the atoms can overcome higher energy barriers and then slowly cool down to approach the energy minimum.

In the application of this method to crystallographic refinement, the calculated structure factors of the system are restrained to the observed structure factors as target values, by adding the crystallographic discrepancy term

$$E_x = k_x \times \sum_{hkl} \left[|F_o(hkl)| - k|F_c(hkl)| \right]^2$$

as a pseudoenergy term to the potential energy E_{pot} of the system. k_x is a weighting factor and k is the scale factor between $|F_o(hkl)|$ and $|F_c(hkl)|$ with $|F_c(hkl)|$ calculated for the present model.

The total energy, E_x + E_{pot}, is minimized in the simulated-annealing refinement. In principle $\mathbf{F}_c$ must be calculated after each time step Δt as are the derivatives of $|F_c|$ with respect to atomic parameters. These derivatives are required for calculating the "force". However, the atomic parameters change very little during Δt and therefore, it is sufficient to calculate $\mathbf{F}_c$ and its derivatives until a preset change in the coordinates has been reached, e.g., 1/10 of the resolution.

2.5.3 Powell minimization

Powell minimization is the energy minimization refinement using the conjugate gradient method of Powell (1977). The conjugate gradient method is an iterative process that gives the solution for an n-dimensional problem in

approximately n steps. In the Powell refinement, the minimization is started from the atomic coordinates x, y, z, and the minimized coordinates are returned in x, y, z to start another cycle of minimization. The basic idea of the conjugate gradient method is as following.

Let us assign $\boldsymbol{x}$ as the vector of variables and $F(\boldsymbol{x})$ as the objective function of $\boldsymbol{x}$. The gradient vector $\boldsymbol{g}(\boldsymbol{x})$ can be calculated

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla F(\boldsymbol{x})$$

The conjugate gradient method does not require any second derivatives. It is an iterative process and converges to the point in the spaces of variables at which $F(\boldsymbol{x})$ is least. If k is assigned as the number of the current iteration starting with $k = 1$, the gradient vector required for the iteration process is

$$\boldsymbol{g}_k = \boldsymbol{g}(\boldsymbol{x}_k)$$

For $k = 1$, let $\boldsymbol{d}_k$ be the steepest descent direction

$$\boldsymbol{d}_k = -\boldsymbol{g}_k$$

If $k > 1$, $\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}$ where $\beta_k = \|\boldsymbol{g}_k\|^2 / \|\boldsymbol{g}_{k-1}\|^2$.

The vector $\boldsymbol{x}_{k+1}$ is obtained by searching for the least value of $F(\boldsymbol{x})$ from $\boldsymbol{x}_k$ along the direction $\boldsymbol{d}_k$.

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$$

where λ_k is the value of λ that minimizes the function of one variable

$$\phi_k(\lambda) = F(\boldsymbol{x}_k + \lambda \boldsymbol{d}_k)\,.$$

This function completes the iteration and another cycle is started if $F(\boldsymbol{x}_{k+1})$ or $\boldsymbol{g}_{k+1}$ is not sufficiently small.

Chapter 3. Crystal Structure of RMP at 2.15 Å

3.1 Structure solution and refinement

3.1.1 Purification

The first and the most important step in protein crystallographic studies of proteins is the crystallization of the proteins. The crystallization requires pure protein samples. RMP was obtained from the fermentation of *Rhizomucor miehei*. The purification procedures of RMP were published by Jia *et al.* (1995) in preliminary studies of RMP. The protocols used for purification of this enzyme are as follows:

(1) Ion-exchange chromatography on SP-Sephadex C50 in 0.05 M citrate buffer, pH 3.2, with 0.1 M sodium chloride and elution by 0.1-0.3 M sodium chloride gradient.

(2) Concentration of the fractions with milk-clotting activity by ultrafiltration using DDS membrane type 800 (M_r cutoff is about 8 to 10 kDa).

(3) Exchange of buffer on a Sephadex G50 medium eluted with 0.02 M piperazine/HCl buffer, pH 5.3.

(4) Ion-exchange chromatography on DEAE-Sephadex A50 in 0.02 M piperazine/HCl buffer, pH 5.3, and elution by a 0–0.6 M sodium chloride gradient.

(5) Concentration by ultrafiltration using a DDS membrane type 800.

(6) Gel-filtration on Sephadex G75 superfine using 0.1 M acetate buffer, pH 6.0, with 0.1 M sodium chloride.

(7) Concentration by ultrafiltration and continuous washing with distilled water, and then freeze drying.

In this research project, 500 mg RMP sample purified by the above protocols was provided by the laboratory of Dr. Palle Schneider, Department of Enzyme Research, Novo Nordisk, DK-2880 Bagsvaerd, Denmark.

3.1.2 Crystallization

The crystallization conditions of RMP were first published by Jia *et al.* (1995). Subsequently better quality crystals were obtained under improved conditions by a repeated seeding method (Thaller *et al.*, 1981). Needle-like

seed crystals were grown by the hanging drop vapour diffusion method with 25 mg/mL protein solution (10 mM sodium citrate buffer, pH 3.6, 6% PEG 8000) suspended over 12% PEG 8000 reservoir solution (10 mM sodium citrate buffer, pH 3.6) at 21°C. Seed crystals were washed by transferring them to 14% PEG 4000 (10 mM sodium citrate buffer, pH 4.0), and put into 5 mg/mL protein drops (5 mM sodium citrate buffer, pH 4.0, 10% PEG 4000) to equilibrate over 13% PEG 4000 reservoir solutions (10 mM sodium citrate buffer, pH 4.0). After two weeks, the seed crystals had grown a little larger. These larger crystals were used as seeds for another cycle of seeding under the same conditions. Large crystals suitable for data collection were obtained after 4 ~ 5 cycles of seeding. Because repeated seeding will often increase the crystal mosaicity, a compromise must be made between the crystal size and the crystal mosaicity. The crystals belong to the orthorhombic space group $P2_12_12_1$. The cell dimensions were: a = 41.86 Å, b = 51.21 Å, c = 174.24 Å, $\alpha = \beta = \gamma = 90°$. There is one enzyme molecule per asymmetric unit. The molecular mass of the protein is 38.7 kDa and the volume per unit mass or Matthews coefficient (Matthews, 1968), V_m, is 2.41 Å^3/dalton. The solvent content is estimated at 49% (Jia *et al.*, 1995). Detailed crystal data are summarized in Table 3.1.

Table 3.1. Crystal data of *Rhizomucor miehei* aspartic proteinase

Molecular mass	38.7 kDa.
Number of residues	361
Crystal size	0.50 × 0.35 × 0.02 mm
Space group	$P2_12_12_1$
Cell dimensions	a = 41.86Å, b = 51.21 Å, c = 174.24 Å
Cell content	One molecule/asymmetric unit, Z = 4
Solvent content	49%

3.1.3 Data collection and processing

The crystals of RMP were mounted in thin-walled glass capillary tubes 1 mm in diameter. X-ray data were collected to a maximum resolution of 2.15 Å at beamline X11 (EMBL beam line), DORIS ring, DESY synchrotron in Hamburg. The wavelength used was 0.927 Å. The oscillation angle was 0.7 degree. The data were collected with a MAR image plate at 20°C and processed using the programs DENZO and SCALEPACK (Otwinowski, 1993) and the CCP4 suite (Collaborative Computational Project #4, 1994). A total of 121,713 reflections to 2.15 Å resolution were merged to yield 19,858 unique reflections with a merging R_{sym} of 0.058 for symmetry-equivalent reflections

based on intensities. The data set was 93.8% complete from 15 to 2.15 Å. The R_{sym} and completeness for each of the resolution shells are summarized in Table 3.2.

3.1.4 Phasing results – Molecular replacement

The crystal structure of RMP was solved by the molecular replacement method. The preliminary molecular replacement solution was obtained by Jia *et al.* (1995) with an incomplete data set of poor quality, using the coordinates of *Mucor pusillus* aspartic proteinase (MPP) as a search model. Some flexible regions of the MPP search model were omitted. The molecular replacement calculations were repeated with 10 to 4.0 Å, 10 to 3.5 Å and 10 to 2.8 Å resolution synchrotron data, respectively, using the program AMoRe (Navaza, 1994); essentially the same result was obtained. In the rotation search, the strongest peak was about 10 σ strong and the second strongest peak was about 5 σ strong. After the translation search the R-factor was 48.0%. The rotation and translation results are summarized in Table 3.3. Rigid body refinement using the program X-PLOR (Brünger, 1992b) was first applied to the entire molecule and then allowing relative movement of the two domains (residues 5–186 and 187–361) in the resolution range 10 to 2.8 Å.

Table 3.2. R_{sym} and data completeness for each resolution shell (I > 0)

Resolution range (Å)	R_{sym}*	Completeness (%)
15.00 ~ 5.73	0.036	97.5
5.73 ~ 4.59	0.039	97.7
4.59 ~ 4.02	0.045	98.3
4.02 ~ 3.66	0.050	98.9
3.66 ~ 3.40	0.053	98.0
3.40 ~ 3.20	0.060	97.1
3.20 ~ 3.05	0.070	97.3
3.05 ~ 2.91	0.078	96.3
2.91 ~ 2.80	0.091	95.2
2.80 ~ 2.71	0.102	94.2
2.71 ~ 2.62	0.114	94.7
2.62 ~ 2.55	0.124	92.4
2.55 ~ 2.48	0.139	93.4
2.48 ~ 2.42	0.157	91.6
2.42 ~ 2.37	0.162	91.5
2.37 ~ 2.32	0.169	91.1
2.32 ~ 2.27	0.182	89.3
2.27 ~ 2.23	0.197	89.9
2.23 ~ 2.19	0.213	86.2
2.19 ~ 2.15	0.261	83.8
All reflections	0.058	93.8

* $R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma \langle I \rangle$

Table 3.3. Molecular replacement result

Rotation (degree)	$\alpha = 115.55$
	$\beta = 69.30$
	$\gamma = 283.97$
Translation (fractions of cell edges)	$\Delta x = 0.290$
	$\Delta y = 0.498$
	$\Delta z = 0.413$

3.1.5 Model building and refinement

Using the two domain rigid body refinement, it was found that the position of the C-terminal domain was statistically significantly different from the position in the preliminary solution of Jia *et al.* (1995). The *R*-factor was 47.7% at this stage. The model was then subjected to further refinement using the molecular dynamics program X-PLOR (Brünger, 1992b). The refinement was carried out using the reflections with $F_o > 2\sigma(F_o)$ in the resolution range 6.0 to 2.15 Å. The general refinement protocol was used: (1) check step (to obtain weights for the diffraction data); (2) preparation stage (to relieve strains or bad contacts of the initial coordinates); (3) slow cooling refinement (simulated-annealing refinement); and (4) Individual B-factor

refinement. Powell minimization was used in the preparation and slow cooling stages. In the slow cooling refinement, the temperature was raised to 3000 K first, and then slowly cooled to 300 K in steps of 50 K with energy minimization in each step. During each cycle of refinement, a SIGMAA weighted $2F_o$-F_c map (Read, 1986) was calculated and model adjustment was performed with the use of the TURBO-FRODO program (Roussel *et al.*, 1990) on a Silicon Graphics Indigo II computer. In total, 40 cycles of model building and refinement were performed, yielding a final *R*-factor of 21.5% for 18,842 reflections with F_o greater than $2\sigma(F_o)$ in the range from 6.0 to 2.15 Å. Water molecules with electron density > 3.5 σ in the SIGMAA weighted F_o-F_c map were added to the model in the last seven cycles. After the B-factor refinement of each cycle, water molecules with B > 70 Å^2 were rejected. In the final model, 140 water molecules were included. The "free-*R*-factor" (*R*-free) (Brünger, 1992a) was also calculated for every cycle except the last cycle of refinement by omitting a set of randomly selected data that comprised about 10% of the observed data. The *R*-free in the second last cycle refinement was 28.1%. In the last cycle of refinement, all of the data with $F_o > 2\sigma(F_o)$ were included and final *R*-factor was 21.5%. The geometry of the final model was examined by the geometry analysis protocol (geomanal) in the program X-PLOR. The results of the geometry analysis, together with the

X-PLOR refinement results, are summarized in Table 3.4. The coordinates of RMP have been deposited in the Brookhaven Protein Data Bank with ID code 2ASI.

Table 3.4. X-PLOR refinement and geometry analysis results

Number of protein non-hydrogen atoms	2644
Number of water oxygen atoms	140
Number of carbohydrate non-hydrogen atoms	39
R.m.s. coordinate error (Å) from a Luzzati plot	0.2
Number of unique reflections used in refinement	18,842
Resolution range (Å)	2.15 to 6.0
Conventional *R*-factor	21.5%
R-free	28.1%
R.m.s.d. for bond distances (Å)	0.01
R.m.s.d. for bond angles (degree)	1.7
R.m.s.d. for dihedral angles (degree)	27
R.m.s.d. for impropers (degree)	1.5
Average B-factor for the protein ($Å^2$)	39.1
Average B-factor for N-terminal domain ($Å^2$)	29.1
Average B-factor for C-terminal domain ($Å^2$)	48.0
Average B-factor for water molecules ($Å^2$)	45.9
Average B-factor for carbohydrates ($Å^2$)	63.7
Average B-factor for residues 82-88 ($Å^2$)	66.6

3.2 Results and discussion

3.2.1 Quality of the model

The crystal structure of RMP has been refined to a conventional crystallographic R-factor of 21.5% and a R-free of 28.1% at 2.15 Å resolution. In the final model, 2644 protein non-hydrogen atoms, 140 water molecules and 39 carbohydrate atoms are included. Four N-terminal residues, Gly85 and the side-chains of eighteen surface residues were omitted in the refinement due to insufficient indication of their positions in the final electron density map. Out of the eighteen omitted surface residues, three are in the N-terminal domain and the other fifteen are in the C-terminal domain. These residues are listed in Table 3.5. The final results of the refinement are given in Table 3.4. Electron density maps ($2F_o$-F_c) showing the active site and part of the β-sheet structures are reproduced in Figure 3.1.

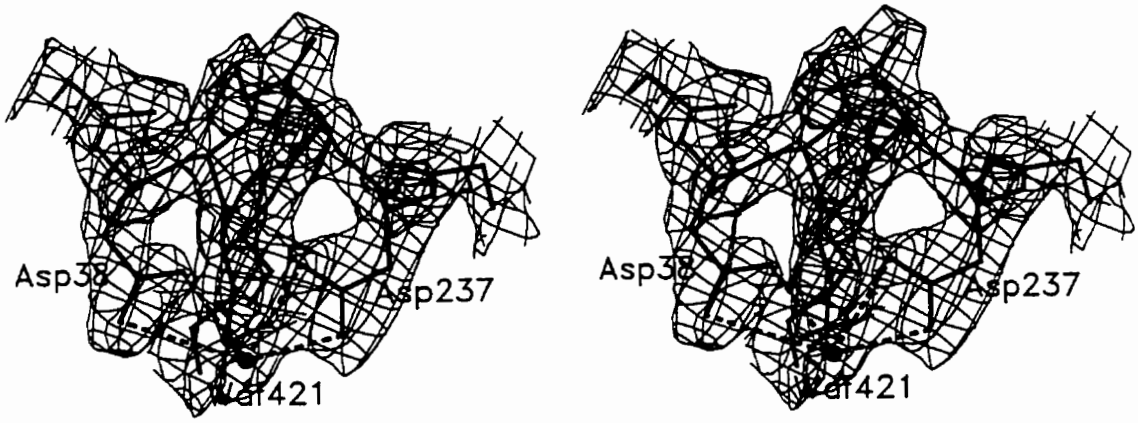
The main chain torsion angles (ϕ, ψ) of RMP were examined by the program PROCHECK (Laskowski *et al.*, 1993). A Ramachandran plot (Ramachandran & Sasisekharan, 1968) is given in Figure 3.2. Except for the glycine and proline residues, all of the residues are in the allowed regions.

Table 3.5. Omitted regions in RMP model

Ala1 to Asp4	All atoms
Lys53	Side-chain atoms
Thr84	Side-chain atoms
Gly85	All atoms
Arg115	Side-chain atoms
Arg204	Side-chain atoms
Ser223	Side-chain atoms
Arg227	Side-chain atoms
Ser248	Side-chain atoms
Ser251	Side-chain atoms
Glu263	Side-chain atoms
Gln266	Side-chain atoms
Ser274-Tyr275	Side-chain atoms
Asn277-Lys279	Side-chain atoms
Glu297	Side-chain atoms
Asp310	Side-chain atoms
Asn326	Side-chain atoms

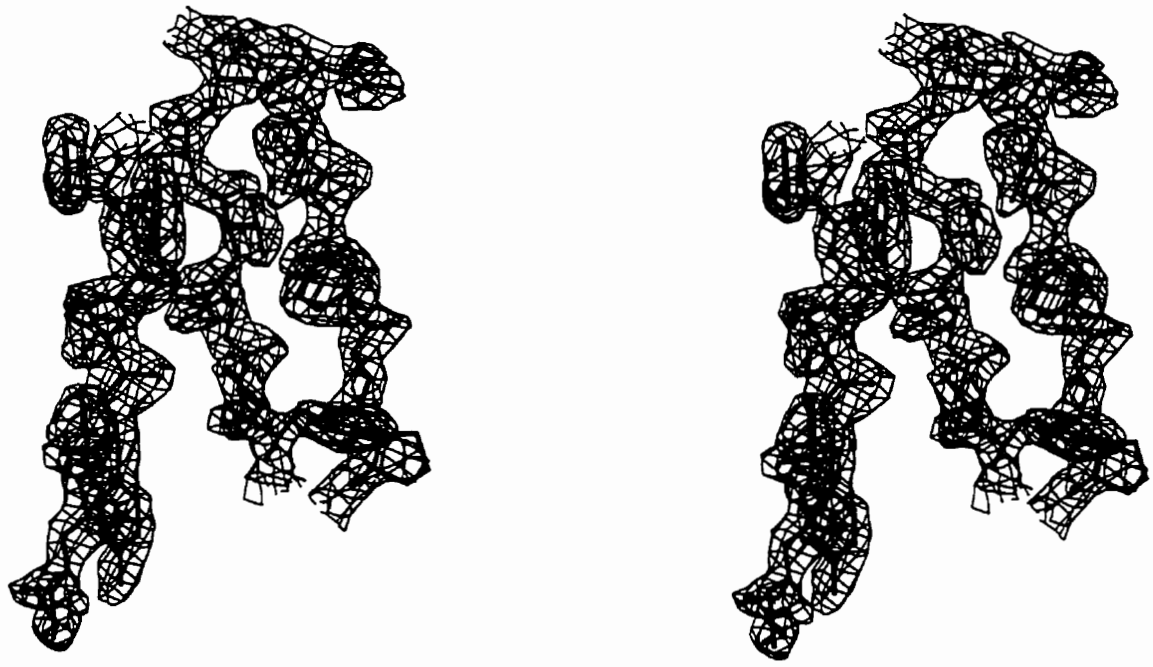
The average B-factors for the main chain and side chain atoms of each residue in RMP have been calculated. The diagrams of the average B-factors of main chain and side chain atoms versus the residue numbers are shown in Figure 3.3. From the figure, it can be seen that the B-factors for the

C-terminal domain residues are higher than those for the N-terminal domain residues. Within the N-terminal domain, residues 82 to 88 show the highest B-factors.



(a). Active site

Figure 3.1. Stereo electron density maps ($2F_o$-F_c) at the active site (a) and part of the β-sheet structures (b). This Figure was prepared by SETOR (Evans, 1993).



(b). Part of the β-sheet structures.

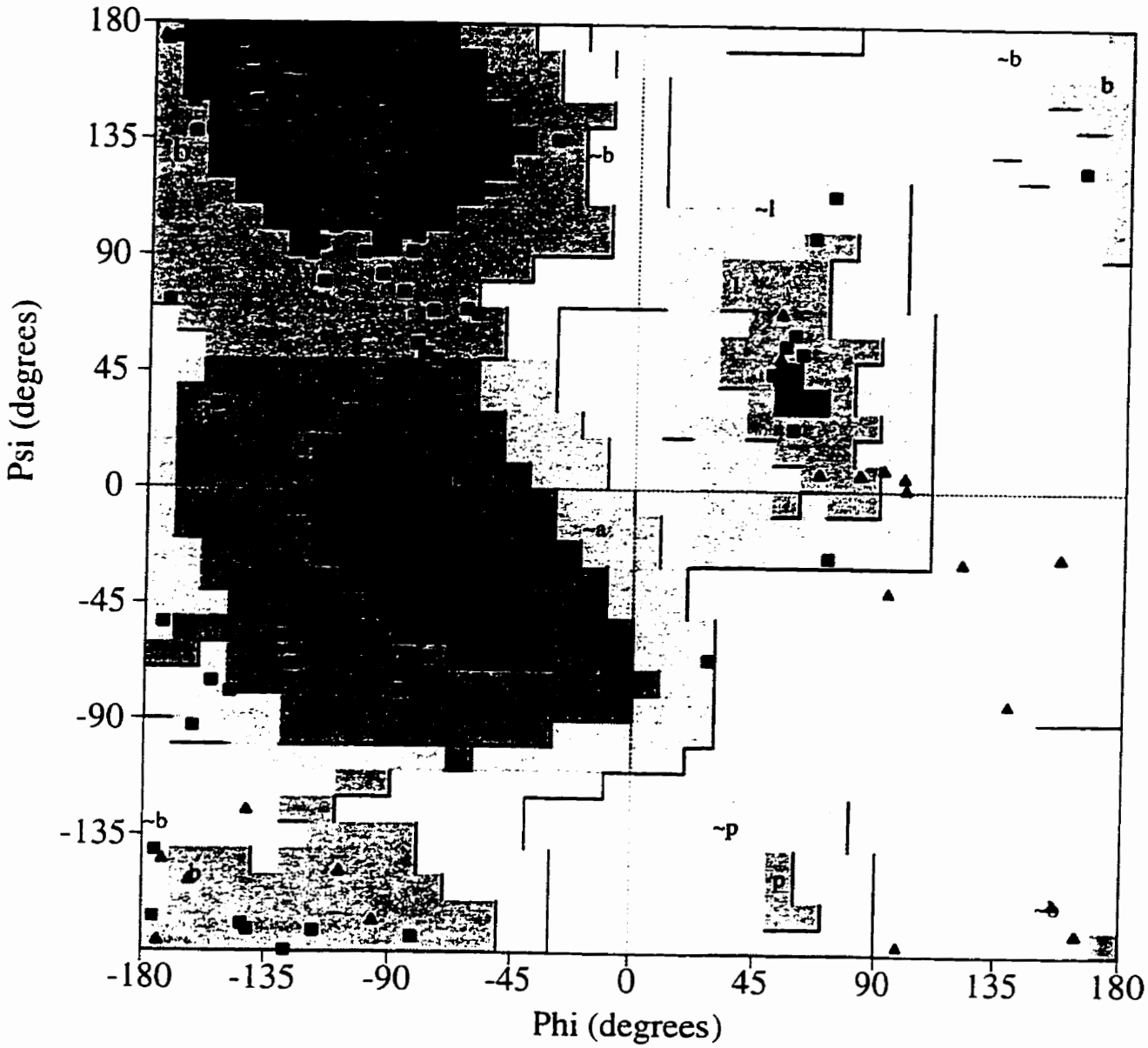


Figure 3.2. Ramachandran plot of RMP. Glycine residues are shown as triangles.

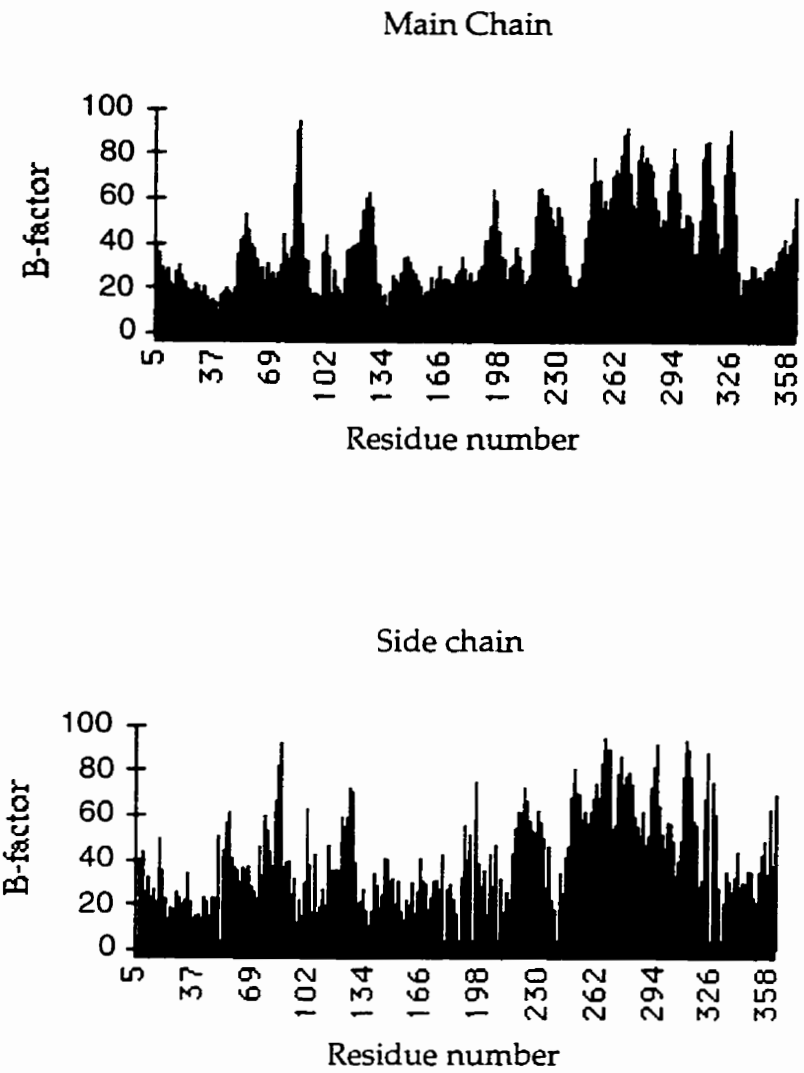


Figure 3.3. B-factors of main chain and side chain atoms versus residue numbers.

3.2.2 Overall fold

RMP contains two domains of approximately the same number of residues, consisting predominantly of β-sheet structures. The N-terminal domain includes residues 1 to 186 and the C-terminal domain includes residues 187 to 361. A ribbon representation of the enzyme secondary structure and overall folding is shown in Figure 3.4. It is obvious that the active site and substrate binding site are located in a deep cleft situated between the two domains. The two catalytic aspartates are situated in the middle of the cleft, with one contributed from each of the domains. Structural studies of the inhibitor-enzyme complexes of other aspartic proteinases (Bott *et al.*, 1982; Suguna *et al.*, 1987; Foundling *et al.*, 1987a,b; Hoover *et al.*, 1991; Pitts *et al.*, 1995) have shown that the inhibitors fit into the cleft upon binding to the enzymes. The average B-factor for the C-terminal domain of RMP is 48.0 $Å^2$, which is 18.9 $Å^2$ higher than the average B-factor of 29.1 $Å^2$ for the N-terminal domain. This indicates that the C-terminal domain is more mobile than the N-terminal domain in the RMP molecule. Šali *et al.* (1989) reported that only a sub-domain in the C-terminal domain, which is about 70% of the C-terminal domain, is flexible in the endothiapepsin molecule. This phenomenon has also been observed in other aspartic proteinases (Sielecki *et al.*, 1990; Cooper *et al.*, 1990; Abad-Zapatero *et al.*, 1991).

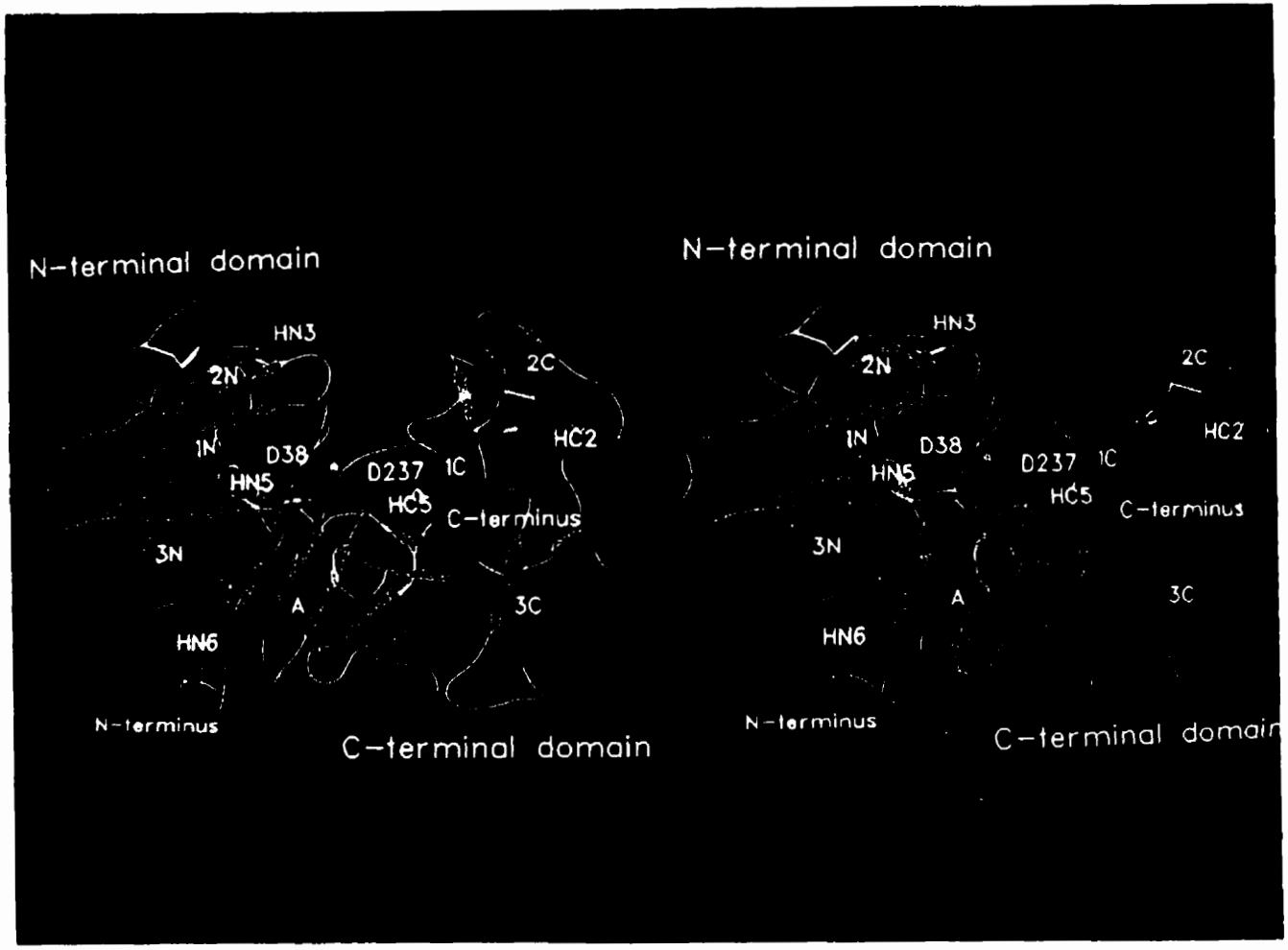


Figure 3.4. The stereo ribbon representation of the secondary structure and overall folding of RMP. This Figure was prepared by SETOR (Evans, 1993).

In order to examine the presence of such a flexible sub-domain in RMP, the average B-factor for the corresponding residues (residues 211 to 331) and the average B-factor for the remainder of the residues (rigid group 1: residues 187 to 210 and 332 to 361) in the C-terminal domain of RMP were calculated. The average B-factor for the sub-domain is 54.8 $Å^2$ and the average B-factor for the rigid group 1 is 34.7 $Å^2$. Thus, RMP also has a flexible sub-domain. The average B-factor for rigid group 1 is still about 19% higher than that of the N-terminal domain. Rigid group 1 is between the N-terminal domain and the sub-domain in the C-terminal domain (Figure 3.5). Some residues of rigid group 1 are involved in the formation of the interface between the N and C-terminal domains.

A view of the α-carbon backbone of the RMP molecule is shown in Figure 3.6, and a diagram showing the packing of RMP molecules within the unit cell is presented in Figure 3.7. Apparently the major reason for the different thermal parameters of the two domains in RMP is crystal packing. A space filling model of RMP is shown in Figure 3.8. The N-terminal domain is shown in blue, the C-terminal domain is shown in green, and the residues which make surface contacts with symmetry-related molecules are shown in red. The N-terminal domain has 30 residues involved in surface contacts with symmetry-related molecules and the C-terminal has only 17 residues involved

in surface contacts with symmetry-related molecules. The N-terminal domain has many more surface contacts with symmetry-related molecules than does the C-terminal domain in the crystals. Therefore, the mobility of the N-terminal domain is restricted by the crystal packing as compared to the C-terminal domain.

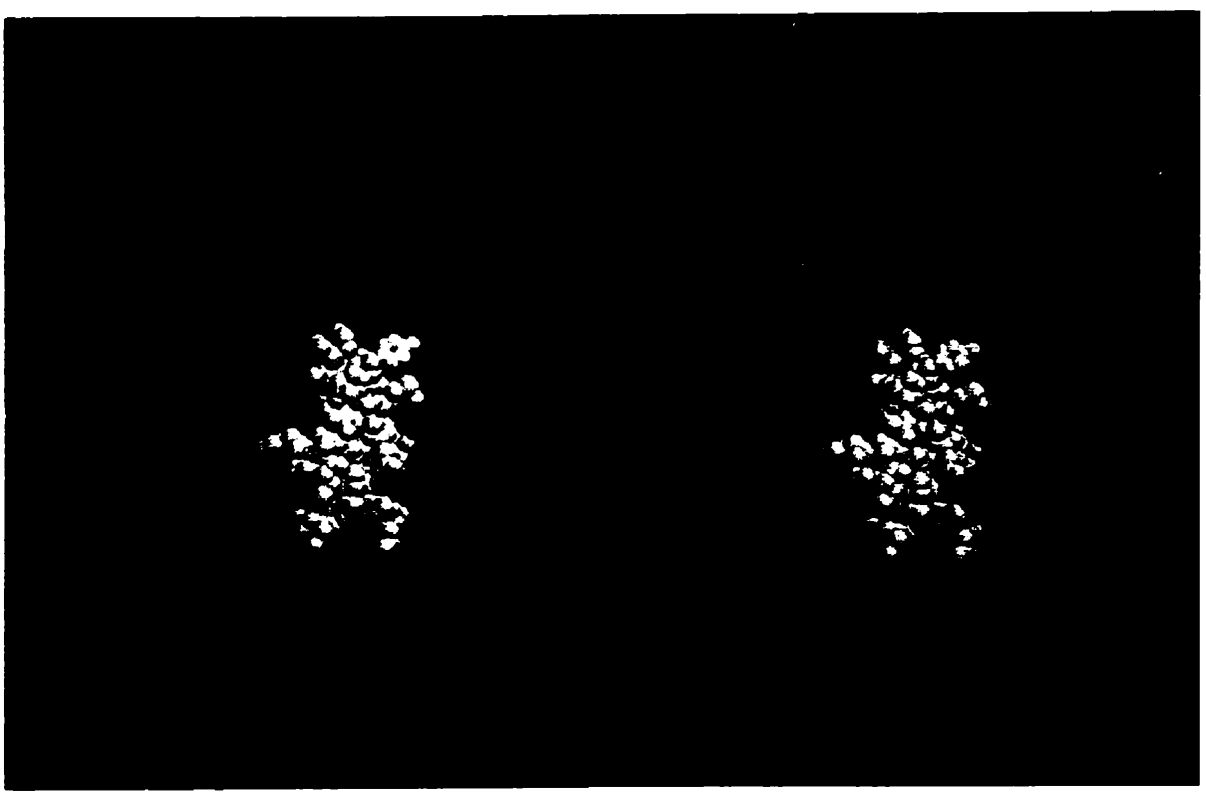


Figure 3.5. A stereo presentation of the N-terminal domain, flexible C-terminal sub-domain and rigid body 1. The N-terminal domain, rigid body 1 and the flexible subdomain are shown in green, yellow and blue, respectively.

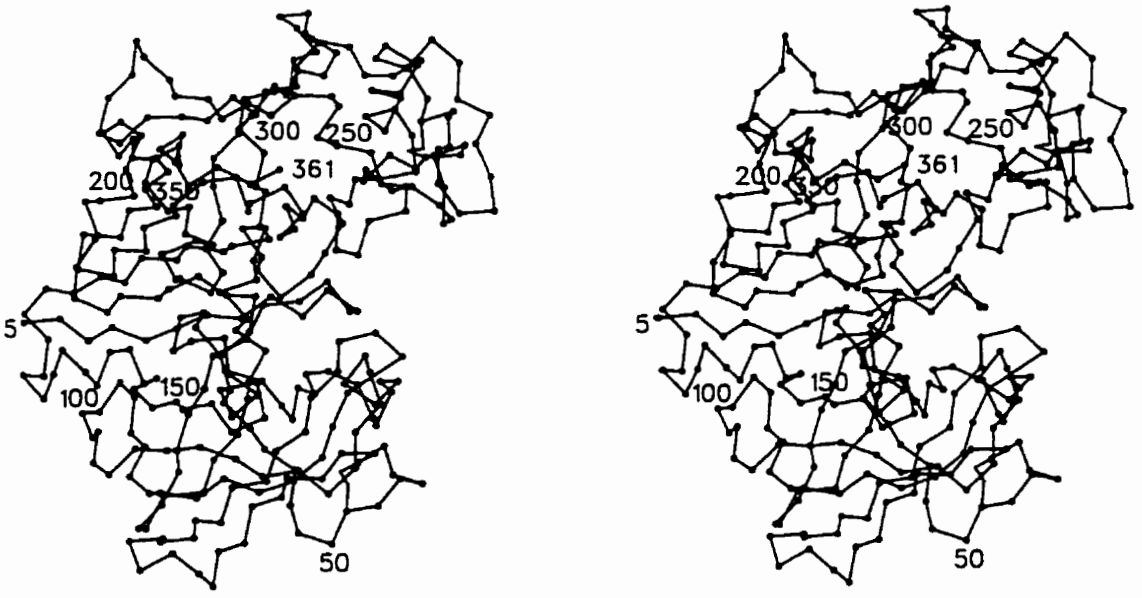


Figure 3.6. Stereo representation of the α-carbon backbone of RMP.

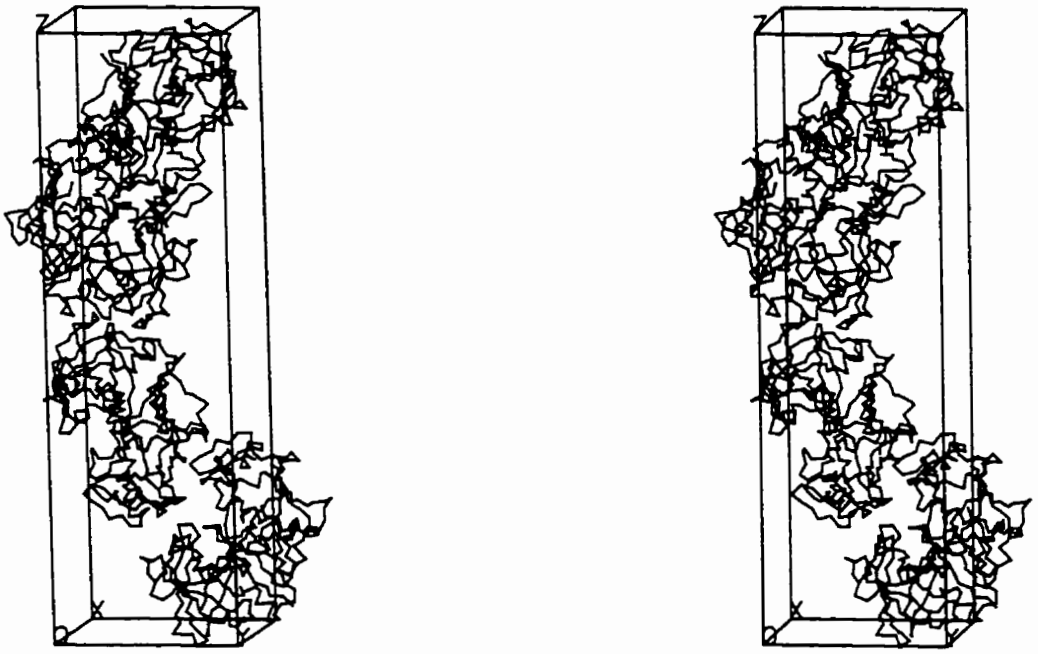
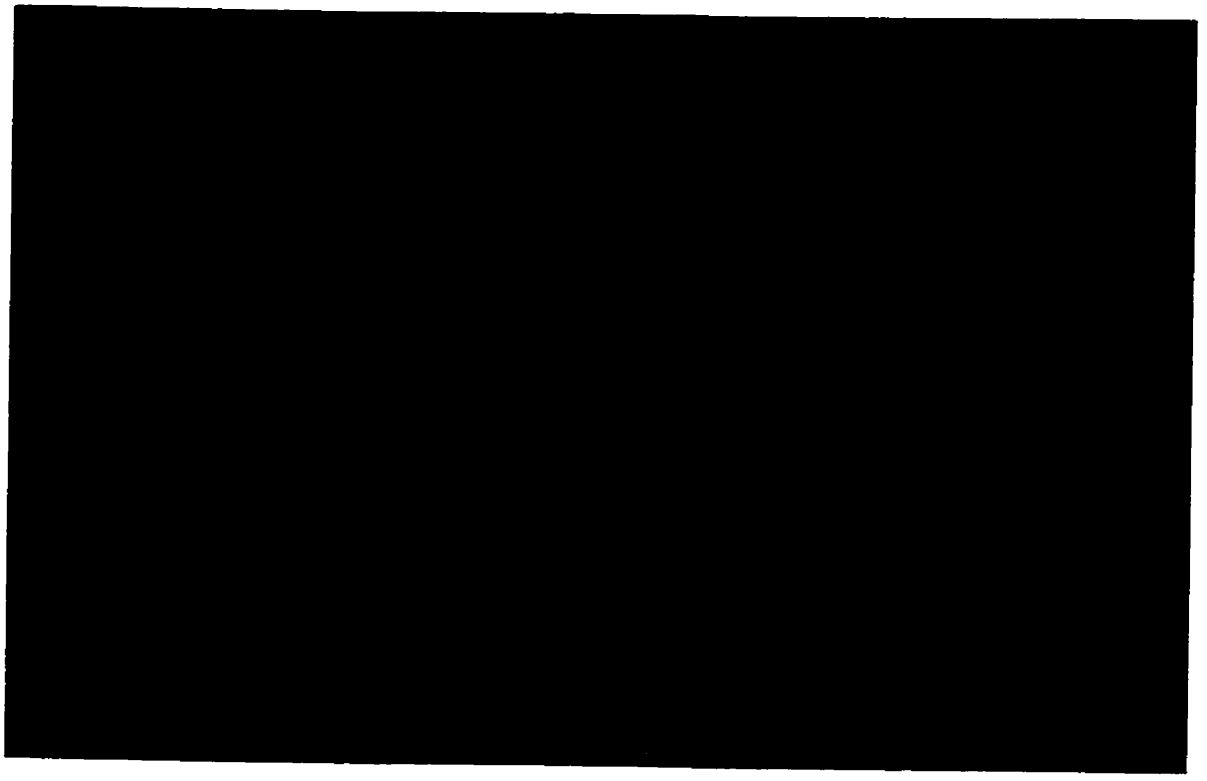


Figure 3.7. The packing of RMP molecules inside the unit cell.



(a). View 1 of Figure 3.8.

(b). View 2.

Figure 3.8. Space filling model of RMP (stereo), showing the interactions with symmetry-related molecules. The N and C-terminal domains are in blue and green, respectively. The residues interacting with symmetry-related molecules are labeled in red. The two views are at approximately 180° to each other.

An active site surface β-hairpin loop (residues 82 to 88) shows the highest temperature factors (average value 66.6 $Å^2$) in the N-terminal domain and weak density in the electron density map ($2F_o$-F_c). Gly85 is omitted in the model because the electron density for this residue is very weak. These observations indicate that this β-hairpin loop is very flexible; it forms a flap over the substrate binding cleft. Crystallographic structural studies of other aspartic proteinases (Bott *et al.*, 1982; James *et al.*, 1982) have shown that this region undergoes a conformational change and becomes more rigid when inhibitors bind to the enzymes. This is due to the formation of hydrogen bonds between the inhibitors and the distal portion of the flap (Bott *et al.*, 1982; James *et al.*, 1982).

There are two disulfide bridges in the enzyme, between cysteines 51 and 57, and between 272 and 316 (Boel *et al.*, 1986) The first disulfide bridge encloses a 3_{10}-helix HN1, while the second links two β-strands, C7 and C11, of the β-sheet 2C. The designations of the secondary structure of RMP will be discussed in section 3.2.3: secondary structure.

3.2.3 Secondary structure

The topological representation of the secondary structure of RMP is shown in Figure 3.9. The secondary structure was calculated by the program PROCHECK (Laskowski *et al.*, 1993) in addition to manual inspection of torsion angles and hydrogen bonding patterns. Both of the two domains consist predominately of β-sheet structure with a few short helical segments.

(i) β-Sheets

The most prominent β-pleated sheet is sheet A, which forms one side of the molecule. It consists of six antiparallel β-strands with three strands from each domain. Sheets 2N, 3N, 2C and 3C are formed of anti-parallel β-strands, with three, four, four and four strands, respectively. Sheets 1N and 1C are formed of mixed β-strands, with eight and five strands, respectively. Sheets 2N, 1N and 3N and sheets 1C and 3C are packed approximately orthogonal in each domain, forming a three and a two-layered sandwich, respectively. This packing has been observed in other proteins having the predominantly β-sheet structure (Chothia & Janin, 1982). The configuration of

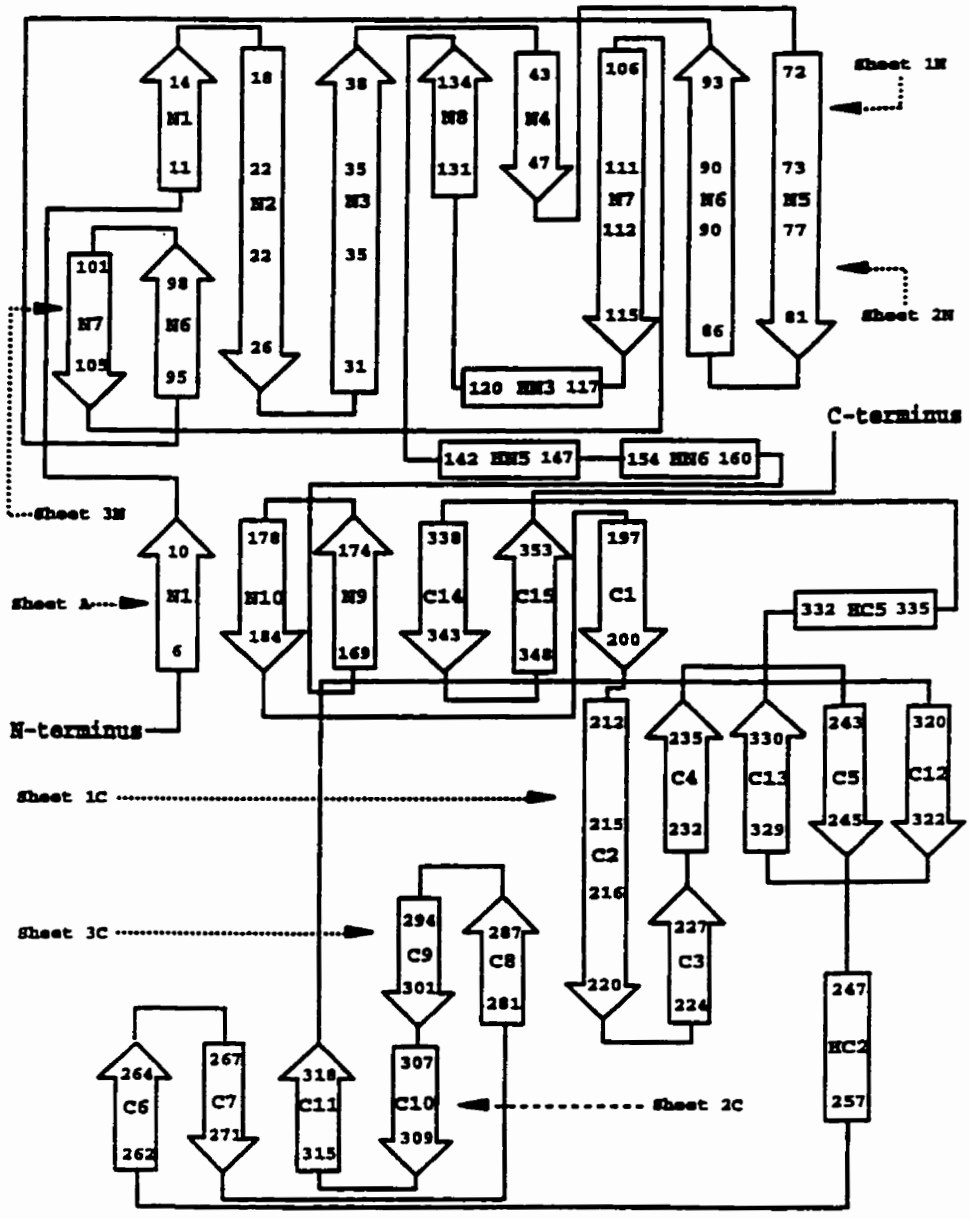


Figure 3.9. Topological representation of the secondary structure of RMP. The β-strands are represented as arrows and the α-helices are represented as rectangles. The 3_{10}-helices are not shown.

three strands, N3, N8 and N4, in sheet 1N resembles the Greek letter ψ, and so this part is also called a ψ-structure. Sheet 2C is at the outer rim of the C-terminal domain. The major force for maintaining this packing is the hydrophobic interactions between the side-chain atoms.

(ii) Helices

There are five α-helices and seven 3_{10}-helices in RMP. The longer helices are all α-helical, whereas the shorter ones are generally 3_{10}-helices. Only the five α-helices are shown in the topological representation of the second structure. The seven 3_{10}-helices are listed in Table 3.6. In contrast to the β-sheets, which are generally conserved, the helices can undergo large variation in positions relative to the β-sheets in different aspartic proteinases.

3.2.4 Active site

In Figure 3.10, a van der Waals surface electrostatic potential model of RMP calculated by the program GRASP (Nicholls *et al.*, 1991) is shown. A large cleft, where the active site and the substrate-binding site are located, is clearly visible between the two domains. The substrate-binding site can

Table 3.6. The 3_{10}-helices in RMP

Name of helices	Initial residue	Terminal residue
HN1	Lys53	Gly56
HN2	Pro65	Ala67
HN4	Pro138	Thr141
HC1	Asn189	Leu191
HC3	Ser274	Gln276
HC4	Asn302	Leu305
HC6	Ser356	Tyr358

accommodate up to eight amino acid residues. The active site has a large negative electrostatic potential. As shown in the ribbon representation of RMP in Figure 3.4, the two catalytic aspartate residues, Asp38 and Asp237, are located at the tips of the ψ-structure and sheet 1C, respectively. For clarity, the residues at the active site are shown in Figure 3.11. The two catalytic aspartate residues are approximately co-planar. The conserved water molecule, Wat421, bridges the two aspartates by forming hydrogen bonds with them at the active site. This type of active site is highly conserved in all aspartic proteinases. Structural analysis of the aspartic proteinases by both X-ray crystallography and computer graphics has shown that the two

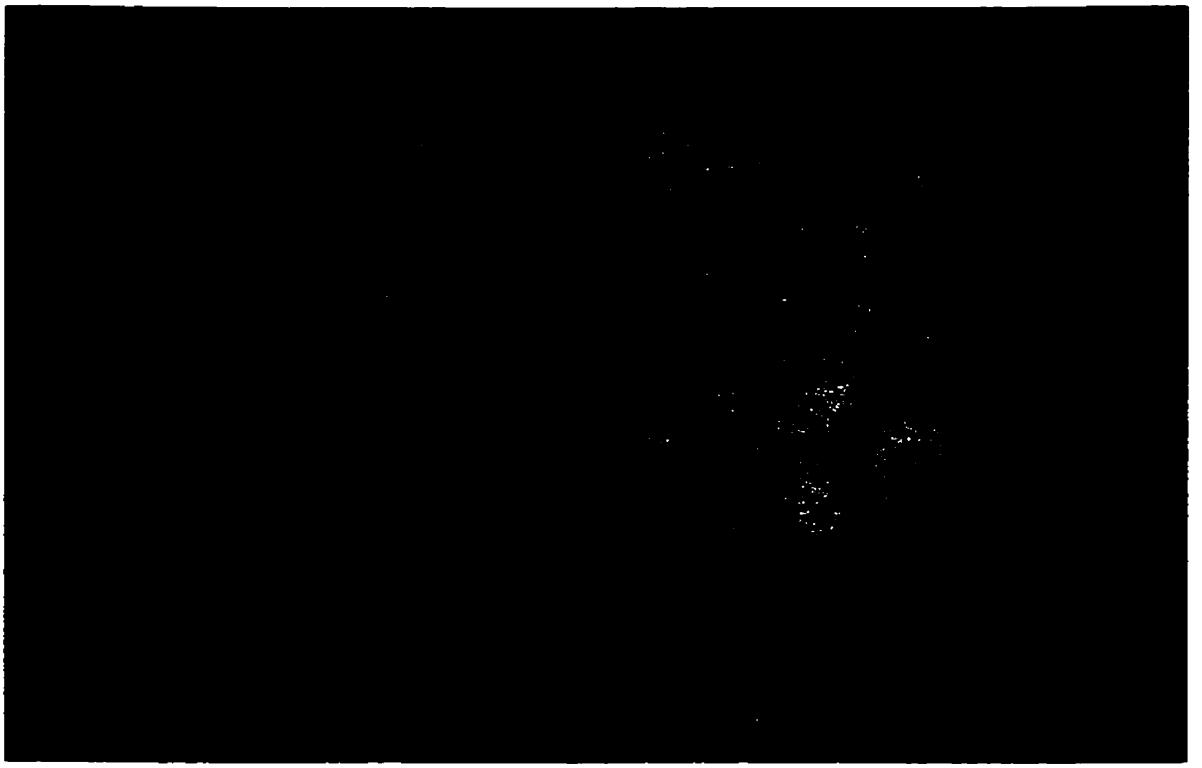


Figure 3.10. Electrostatic potential model of RMP prepared by the program GRASP (Nicholls *et al.*, 1991). Blue and red represent positive and negative electrostatic potentials, respectively.

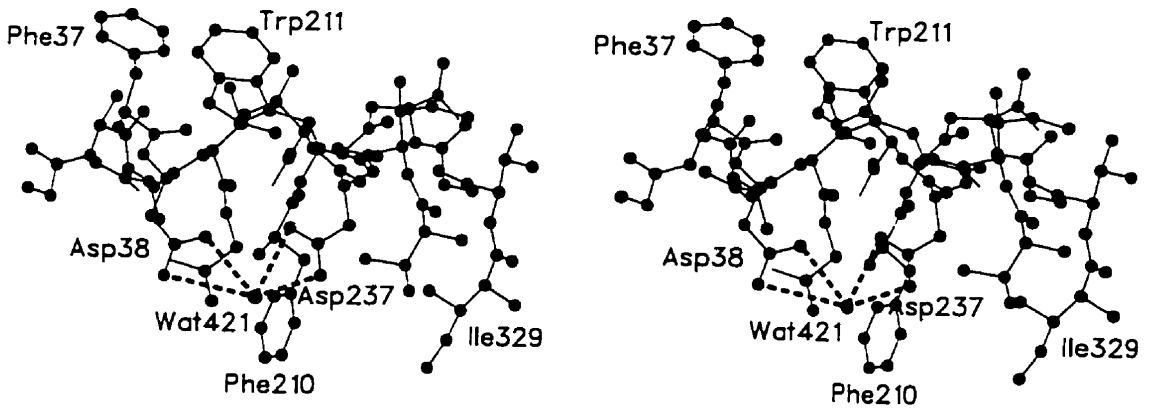


Figure 3.11. The active site of RMP (stereo).

catalytic aspartates are related by a local 2-fold symmetry (Mantafounis & Pitts, 1990). The structure of the ground state for the active site shows that there is probably one proton shared between the carboxylates to produce a net negative charge (Mantafounis & Pitts, 1990). There are quite a few polar residues within 10 Å from the active site. These residues, together with the two catalytic aspartates, are involved in a network of hydrogen bonds, which makes this region relatively rigid. The sequence Asp-Thr-Gly-(Ser/Thr) plays a key role in this network (Newman *et al.*, 1993). The hydrogen bonds in this network are summarized in Table 3.7.

At the time when the RMP native structure was solved, no RMP-inhibitor complex structures had yet been determined. In order to obtain detailed knowledge of the mode of substrate binding in the active site, molecular modeling was used to dock the inhibitor pepstatin A molecule into the active site of RMP. The pepstatin A sequence mimics residues 102 to 108 (His-Leu-Ser-Phe-Met-Ala-Ile) of κ-casein. Pepstatin A coordinates were obtained from the endothiapepsin-pepstatin A complex structure (Bailey *et al.*, 1993). The endothiapepsin-pepstatin A complex molecule was superimposed on the RMP molecule by least squares. The pepstatin A coordinates after superposition were combined with the RMP coordinates to give the RMP-pepstatin A complex. The RMP-pepstatin A complex structure was then

Table 3.7. The network of hydrogen bonds near the active center of RMP

Asp38 OD1	Asp237 OD1	3.10 Å
Asp38 OD2	Wat421	3.47 Å
Asp237 OD1	Wat421	2.95 Å
Wat421	Wat509	3.11Å
Gly40 N	Asp38 OD1	2.72 Å
Ala135 N	Asp38 O	3.12 Å
Thr238 O	Wat420	2.82 Å
Asn332 ND2	Thr238 O	3.20 Å
Asn332 N	Thr240 O	2.75 Å
Gly40 O	Wat473	3.10 Å
Leu36 O	Asp38 N	3.13 Å
Asp38 O	Gly40 N	2.97 Å
Val330 N	Thr235 O	2.83 Å
Asp237 N	Val330 O	3.08 Å
Thr39 N	Thr238 OG1	2.87 Å
Asp38 OD1	Wat421	2.95 Å
Asp237 OD2	Wat421	2.56 Å
Ser41 OG	Asp38 OD2	2.72 Å
Asp38 O	Ser41 N	3.25 Å
Ala135 N	Ser41 O	2.93 Å

Table 3.7 continued.

Thr238 O	Thr240 N	3.21 Å
Asn332 ND2	Thr240 O	3.03 Å
Gly239 N	Asp237 OD1	2.84 Å
Phe133 N	Leu36 O	2.95 Å
Phe133 O	Asp38 N	2.92 Å
Phe37 O	Wat420	3.14 Å
Ile236 N	Trp211 O	2.88 Å
Asp237 N	Thr235 O	3.19 Å
Thr39 OG1	Thr238 OG1	3.21 Å
Gly40 N	Asp237 OD1	3.31 Å
Asp38 O	Ser41 O	3.20 Å
Ala136 N	Ser41 O	2.79 Å
Ser42 N	Wat406	2.72 Å
Thr39 O	Wat404	2.83 Å
Thr235 OG1	Asp212 OD1	2.77 Å
Ile236 O	Wat404	3.14 Å
Thr39 OG1	Thr238 N	2.94 Å
Glu19 OE1	Asn332 ND2	3.30 Å
Glu18 O	Asn332 ND2	3.03 Å
Asn241 ND2	Glu18 OE1	3.15 Å
Ser41 OG	Wat453	2.65 Å
Trp211 N	Wat404	2.95 Å

refined by energy-minimization with the program X-PLOR (Brünger, 1992b). In the energy-minimization process, the temperature was raised to 2000K and then slowly cooled down to 300K. Movement was allowed for both the RMP and pepstatin A molecules. The final structure of the complex (Figure 3.12) shows that the inhibitor pepstatin A fits into the substrate binding cleft in the extended conformation. The possible residues involved in the substrate binding are summarized in Table 3.8. The hydroxyl group at the C3 position of the P1-P1' statine residue forms hydrogen bonds with both catalytic aspartate residues. This conformation is believed to mimic the transition state of the aspartic proteinase-substrate interaction. This structural analysis of the native RMP structure and the RMP-pepstatin A complex model from molecular modeling fully supports the catalytic mechanism proposed for the aspartic proteinases (Pearl, 1987; Suguna *et al.*, 1987; Hoover *et al.*, 1991). The proposed catalytic mechanism is shown in Figure 3.13. About one and a half years after the RMP native structure was solved, the RMP-pepstatin A complex structure was solved at 2.7 Å. This structure supports the modeling studies and the proposed catalytic mechanism. In Chapter 4 of this thesis, the RMP-pepstatin A complex structure and its comparison with the structure obtained by molecular modeling will be discussed.

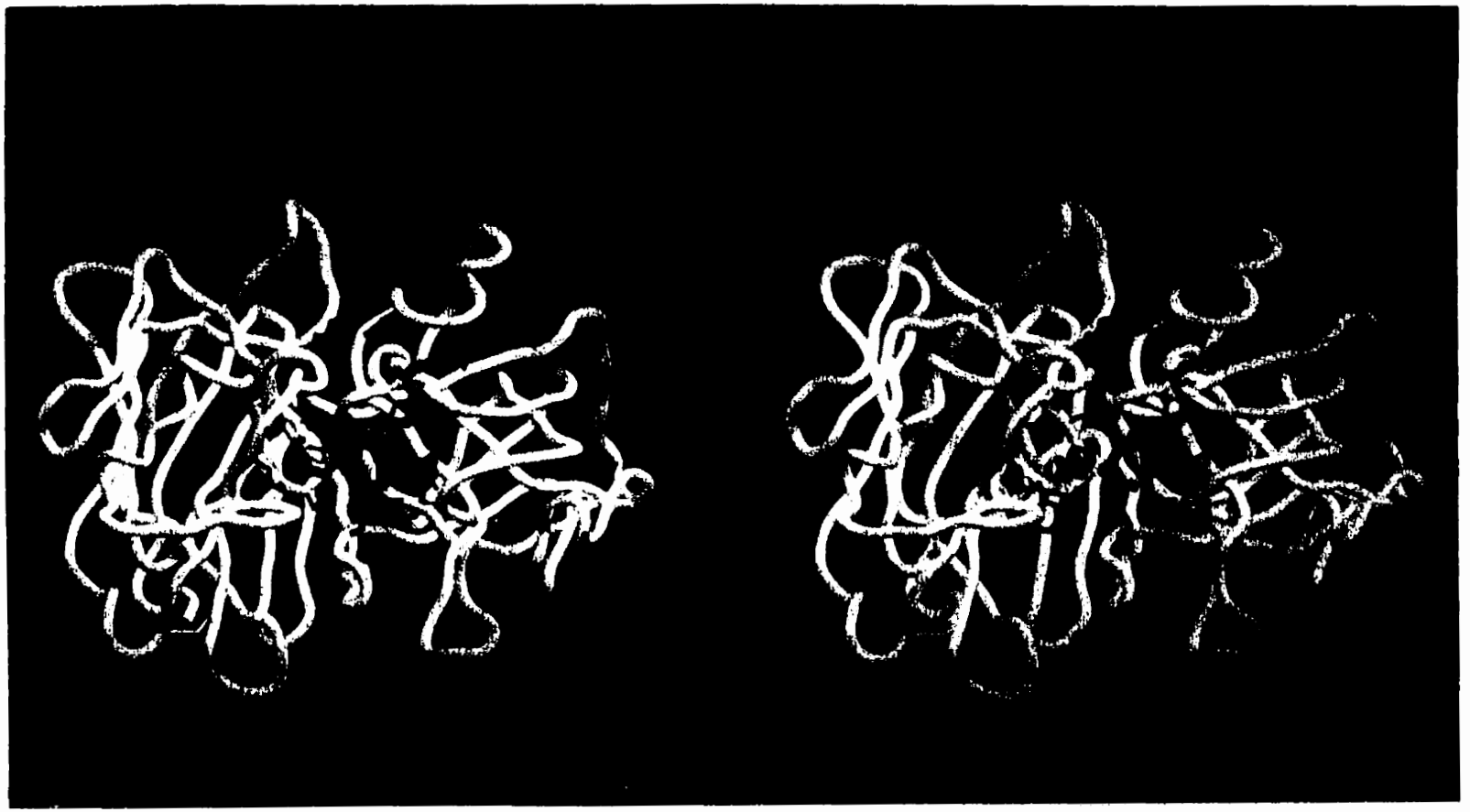


Figure 3.12. The RMP-pepstatin A complex model from molecular modeling (stereo view). The pepstatin A molecules is shown in red. The possible substrate binding residues of RMP are shown in blue.

Table 3.8. Substrate binding site residues in RMP.

Substrate	Component	Subsite residues
P4	IVA	Asn241, Phe242, Ile244, Leu321
P3	Val	Thr240, Asn241, Glu19, Leu17
P2	Val	Thr240, Ile244, Gly325, Ile329, Gly83, Thr84
P1-P1'	STA	Asp38, Asp237, Tyr82, Gly83, Gly329, Thr240, Gly40, Leu36, Leu132, Pro117
P2'	Ala	Tyr82, Ser41, Gly40, Phe210
P3'-P4'	STA	Tyr82, Thr81, Ile80, Phe210, Asn140

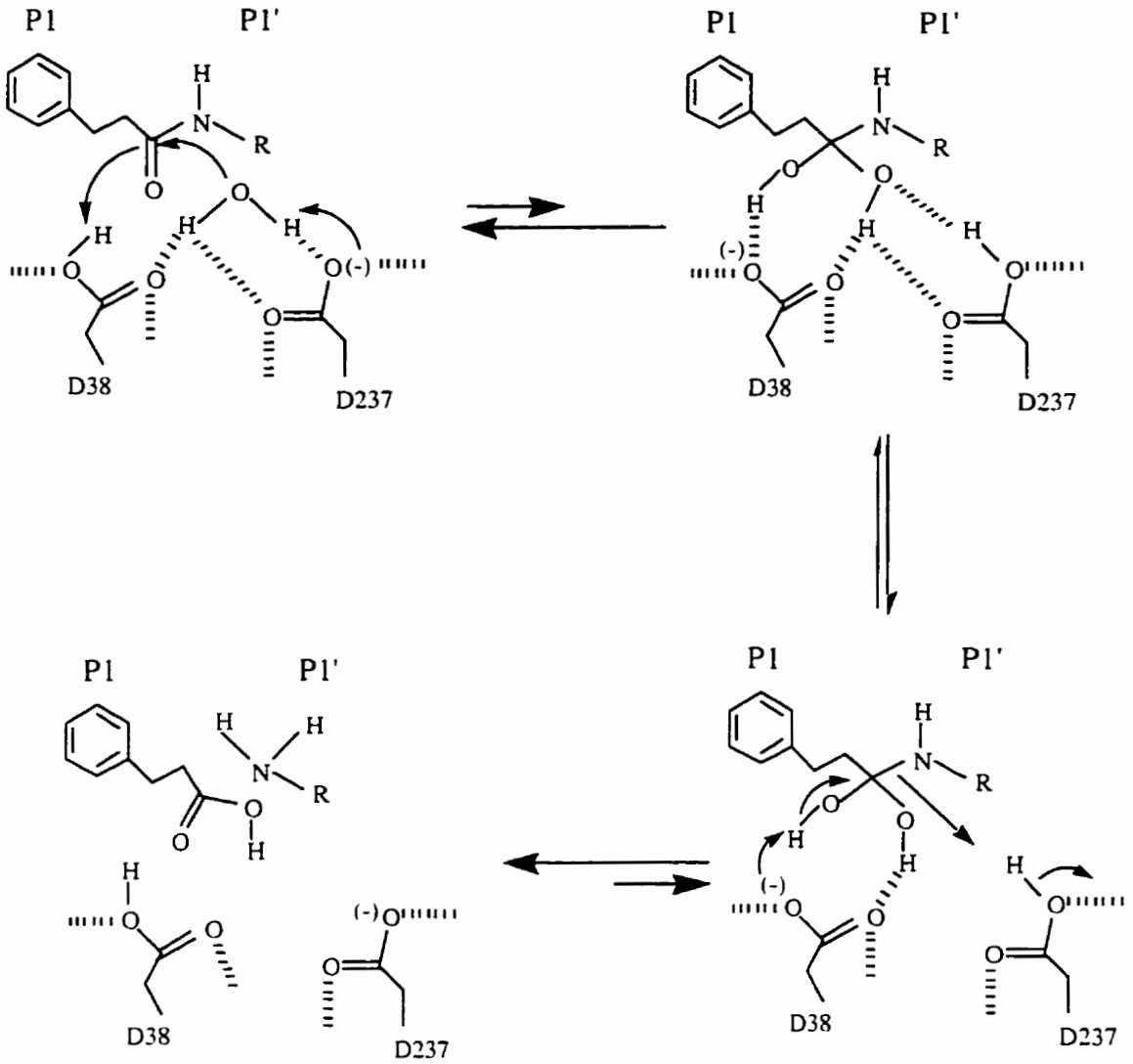


Figure 3.13. The catalytic mechanism of aspartic proteinases (Suguna *et al.*, 1987).

3.2.5 Electrostatic potential and the optimum pH

All of the aspartic proteinases except renin show maximum activity at acidic pH (i.e. pH < 5.0). Renin is usually optimally active at neutral pH (i.e. 6.5 < pH < 7.0). To carry out the functions at low pH, the enzymes must be stable in the acidic environment. Andreeva & James (1991) reported that pepsin has a high ratio of potentially negatively-charged groups to positively-charged groups, and this high ratio gives the enzyme an unusually stable native structure at low pH values. Pepsin has several acidic residues inside the protein with depressed pK_a values. These residues cause pepsin to remain slightly negatively charged even at very low pH values. For most proteins the ratio of negatively-charged groups to positively-charged groups at neutral pH is close to 1.0. Thus, as the net charge on the protein increases at extreme pH values, the charge repulsion increases and destabilizes the protein folding.

To compare different aspartic proteinases, porcine pepsin numbering (Sielecki *et al.*, 1990) will be used. The two catalytic aspartic acid residues in this numbering system are Asp32 and Asp215. The activity at acidic pH for aspartic proteinases has attracted the attention of scientists for many years,

but the reasons for it are still not clear. Studies from two research groups (Yamauchi *et al.*, 1988; Mantafounis & Pitts, 1990) have focused attention on a single residue, residue 303. Most aspartic proteinases (e.g. pepsin, chymosin b) have aspartate at residue 303. Exceptions are renin (alanine), MPP (asparagine) and RMP (asparagine). A similar hypothesis has been proposed from these two groups and is shown in Figure 3.14. They propose that changes in residue 303 contribute to the differences in the pH optimum of catalysis through a mechanism that involves the degree of protonation of the active site aspartate residues being affected by polarization of peptide bond Thr216–Gly217 by Asp303 (protonated) through a hydrogen bond (OD_{303}-$H...O_{216}$). The polarization effect of this peptide bond is transferred to the carboxyl group of Asp215 *via* another hydrogen bond between the main-chain N atom of residue Gly217 and one carboxyl oxygen atom of Asp215 to affect the degree of protonation of Asp215; no such hydrogen bond network can form if residue 303 is alanine. Newman *et al.* (1993) proposed that for MPP the presence of asparagine at residue 303 in place of the usual aspartate results in a more basic pH optimum because of the reduced ability of asparagine to polarize peptide bond Thr216-Gly217 through a hydrogen bond (ND_{303}-$H...O_{216}$). In RMP, the corresponding residue at this position (residue 332) is also asparagine and the pH optimum is 4.7 (Martin *et al.*, 1980), compared to 4.7 for MPP.

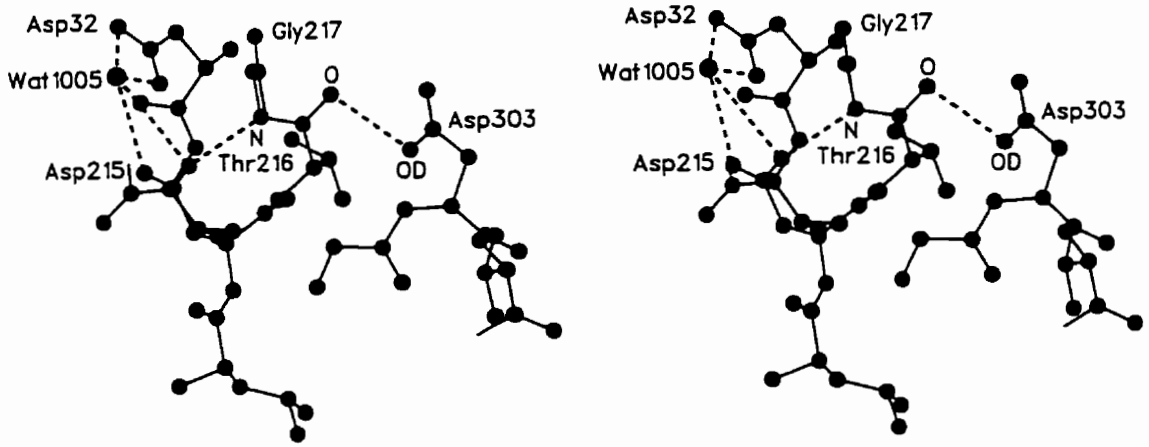


Figure 3.14. The hypothesis proposed to explain why aspartic proteinases are optimally active at acid pH (Yamauchi *et al.*, 1988; Mantafounis & Pitts, 1990).

The hypothesis of Newman *et al.* (1993) regarding the significance of residue 303 MPP was examined and it was concluded that this hypothesis is unlikely to be valid. The energy for a hydrogen bond with an uncharged group in proteins is only about 0.5-1.5 kcal/mol (Fersht *et al.*, 1985). Since the bond energy difference between the hydrogen bond of O–H...O (chymosin B) and N–H...O (MPP) is likely very small (less than 0.5 kcal/mol), the difference in

polarization of the peptide bond Thr216-Gly217 caused by switching one of the oxygen atoms in the hydrogen bond to a nitrogen atom would have little significance. It is unlikely that this small polarization difference would significantly change the ionization state of Asp215 between chymosin B and MPP.

To determine why most aspartic proteinases are most active at acidic pH values, the structures of three enzymes, chymosin B (optimum pH 3.7), RMP (optimum pH 4.7) and renin (optimum pH 7.0), which possess significantly different pH optima for activity, were examined. The ratios of potentially negatively-charged groups to positively-charged groups are 1.6, 1.9 and 1.2 for chymosin B, RMP and renin, respectively. The calculation of this ratio, proposed by Andreeva and James (1991), includes the potentially charged residues both on the surface of a protein and inside the protein. However, several acidic residues are buried beneath the protein surface in the three enzymes. The pH effects on these buried residues are not as significant as on the surface residues because the internal acid residues are not likely to be ionized. Therefore, one would expect that the high ratio of potentially negatively-charged groups to positively-charged groups on the surface of the proteins may be more important in stabilizing the proteins at low pH. The ratios of the surface charged groups are 1.5, 1.6 and 1.0, respectively, for

chymosin B, RMP and renin. These indicate that chymosin B and RMP are stable at low pH values while renin is stable only at pH near 7.0.

There are three acidic residues, six polar residues and no basic residues within 10 Å of the active site in chymosin B; three acidic residues, nine polar residues and one basic residue are within 10 Å of the active site in RMP; there are twelve polar residues and no acidic or basic residues are within 10 Å of the active site in renin. The active site of each aspartic proteinase has a hydrogen bonding network and all of these acidic, basic and polar residues, together with the two catalytic aspartate residues, are involved in the network. Therefore, the influence of these acidic or basic residues on the two catalytic aspartate residues is significantly affected by other residues in the hydrogen bonding network. To correlate the effect of all the residues close to the active site on the catalytic aspartate residues, the surface electrostatic potentials of chymosin B, RMP and renin were calculated by the program DelPhi (Sharp & Nicholls, 1989). The results of the calculations (Table 3.9) show chymosin B and renin have the most and least negative electrostatic potentials at the active sites, respectively. Because the negative electrostatic potentials will decrease the pK_a values of the two aspartate residues and the optimum pH of each aspartic proteinase is approximately the average of the pK_a values of the two aspartate residues, it is proposed that the

Table 3.9. Electrostatic potentials at the active site of chymosin B, RMP and renin

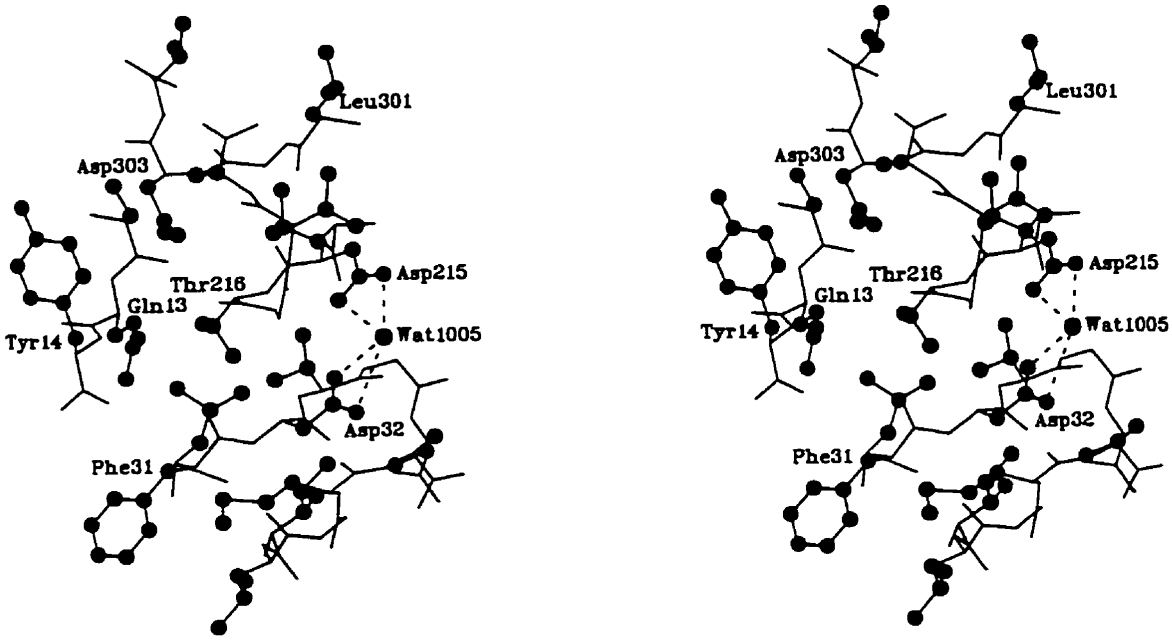
Enzyme	Residue	Atom type	Electrostatic potential (kT)
chymosin B	Asp32	OD1	-650.1
		OD2	-582.5
		Average	-616.3
	Asp215	OD1	-679.1
		OD2	-622.3
		Average	-650.7
D303A chymosin B	Asp32	OD1	-642.2
		OD2	-578.9
		Average	-610.6
	Asp215	OD1	-666.2
		OD2	-616.3
		Average	-641.3
RMP	Asp38	OD1	-683.9
		OD2	-592.8
		Average	-638.4
	Asp237	OD1	-589.0
		OD2	-422.5
		Average	-505.8

Table 3.9 continued.

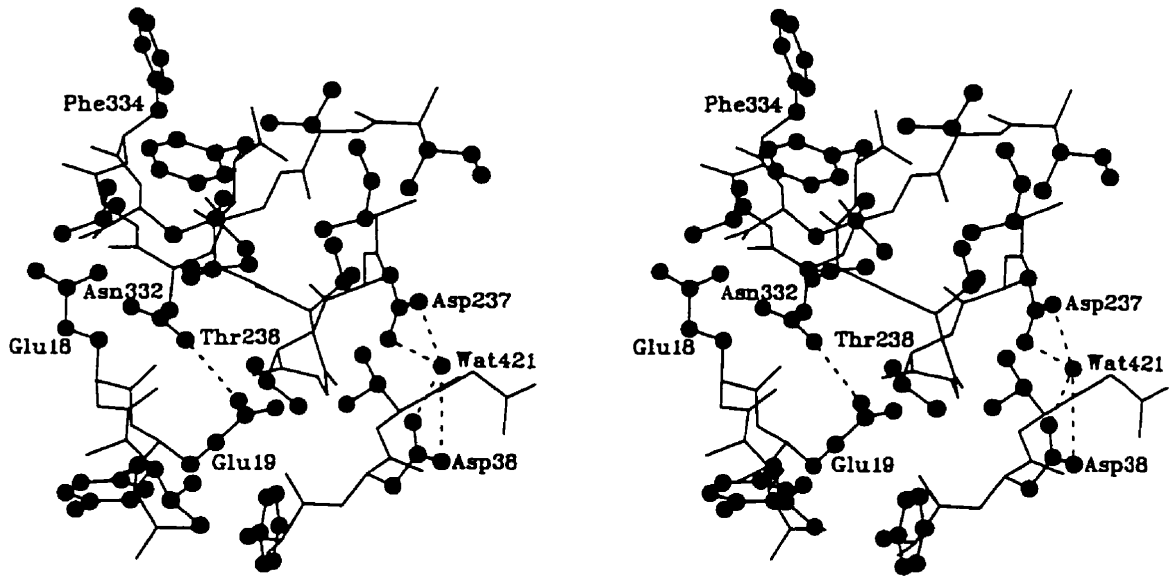
E19A RMP	Asp38	OD1	-673.1
		OD2	-586.6
		Average	-629.9
	Asp237	OD1	-576.0
		OD2	-416.4
		Average	-496.2
renin	Asp38	OD1	-541.4
		OD2	-495.0
		Average	-518.2
	Asp226	OD1	-546.2
		OD2	-628.1
		Average	-587.2
A317D renin	Asp38	OD1	-547.6
		OD2	-497.9
		Average	--522.8
	Asp226	OD1	-555.1
		OD2	-632.9
		Average	-594.0

optimum pH of each aspartic proteinase is determined by the electrostatic potential at the active site. This, in turn, is determined by the positions and orientations of all the residues near the active site.

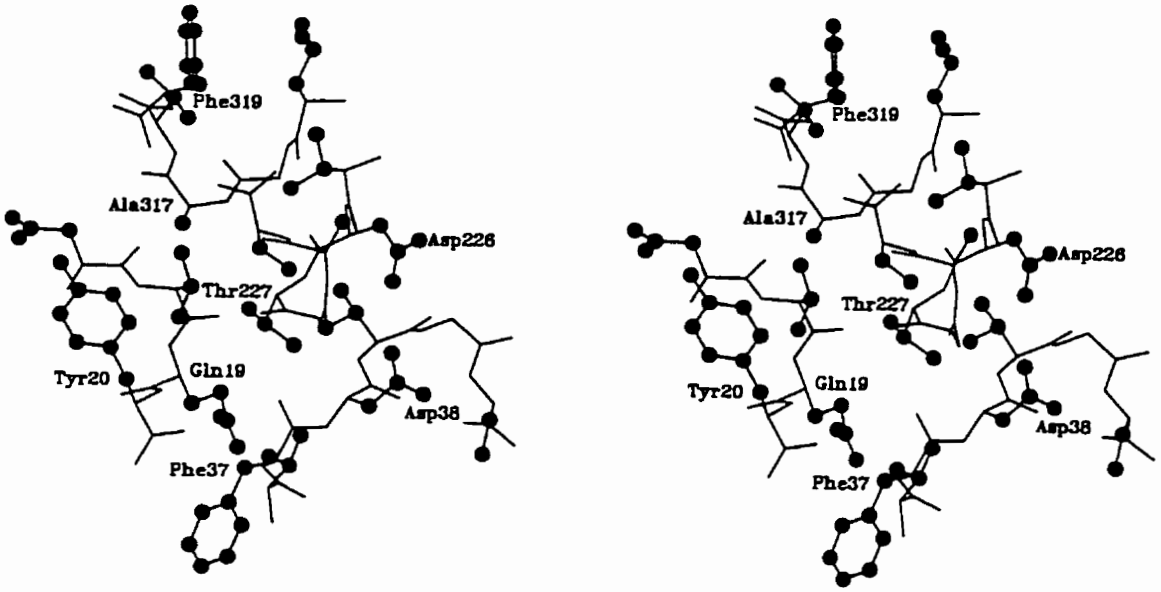
Site-directed mutagenesis studies have shown that the Ala303 to Asp mutation in renin results in a downward shift in the pH optimum of 0.5 (Yamauchi *et al*, 1988) and the change of Asp303 to Ala in chymosin B results in an upward shift of 0.6 unit (Mantafounis & Pitts, 1990). Thus, residue 303 plays an important role in determining the pH optimum of each aspartic proteinase, but other factors must also be important to account for the observed differences. Structural comparisons of chymosin B, RMP and renin (Figure 3.15) show significant conformational differences of residues 303 and 13. In chymosin B, residue 303 is Asp and is the closest acidic residue to the active site (about 6.5 Å). Its side-chain points towards the active site, while residue 13, which is glutamine, is directed toward the outside of the active site and does not interact with Asp303. Asp303 would contribute any negative charge to the active site. In RMP, the residues that correspond to porcine pepsin positions 303 and 13 are Asn332 and Glu19. The side-chain of Glu19 points toward the active site and forms a hydrogen bond with Asn332 (unlike chymosin B, where Gln13 directed away from the active site). In RMP Glu19 is the closest acidic residue to the active site (about 6.5 Å). In renin, the



(a) chymosin B



(b) RMP



(c) renin

Figure 3.15. Stereo diagrams of the active sites of (a) chymosin B, (b) RMP and (c) renin, showing the different conformations of residues 303 and 13 in the three enzymes. This Figure is prepared by SETOR (Evans, 1993).

corresponding residues are Ala317 and Gln19. Gln19 is directed away from the active site in renin.

To isolate the contributions of residue 303 for chymosin B and renin, the electrostatic potentials of the hypothetical mutants Asp303Ala (chymosin B) and Ala317Asp (renin) were calculated. The contribution of Glu19 in RMP

was isolated by calculating the electrostatic potential of the mutant Glu19Ala (RMP). The results (Table 3.9) show that the Asp303 to Ala mutation in chymosin B or the Ala317 to Asp mutation in renin does not cause the electrostatic potentials at the active site to increase or decrease to the level in RMP. These calculations are consistent with the mutagenesis results quoted above. The results also show that changing Glu19 to Ala in RMP does not cause the active site electrostatic potential to the level in renin. Therefore, it is not expected that the change in the pH optimum for RMP is very large although the mutation of Glu19 to Ala might shift the pH optimum upward by approximately 0.5 unit in RMP. The above calculations indicate that the collective effect of all the residues in the region of the active site is important in determining the electrostatic potential of an aspartic proteinase although residues 303 and 13 may play a significant role in determining the electrostatic potentials at the active site. The results are consistent with the present hypothesis in explaining why most aspartic proteinases reach their maximum activity at acidic pH.

3.2.6 Glycosylation and thermal stability

RMP is the most highly glycosylated among the aspartic proteinases, possessing about 6% carbohydrate (Rickert & McBride-Warren, 1974). Previous work (Boel *et al.*, 1986) had shown that carbohydrate units were attached at residues Asn79 and Asn188. By examining the SIGMAA weighted $2F_o$-F_c electron density map (Read, 1986), a disaccharide moiety NAG-Man and a monosaccharide NAG moiety attached to residues Asn79 and Asn188, respectively, were located. Brown & Yada (1991), by comparing RMP with its homologues in their kinetic studies, suggested that Asn79 and Asn188 were on β-turns on the surface of the molecule. The attachment at β-turns allows the typically large N-linked structures free rotation, ensuring the protective function of the carbohydrate moiety (Montreuil, 1984). However, the present results show that neither of the carbohydrate moieties is on a β-turn in RMP. Asn79 is in the middle of a β-strand and Asn188 is on the surface loop which connects the two domains. The attachment of carbohydrate to a β-strand may limit the positional freedom of the carbohydrate due to the steric restraint of the β-strand. The significance of this kind of attachment is still not clear. However, because this carbohydrate moiety is close to the flexible active site surface flap region (residue 82-88), it

may mask the flap region from being attacked by other proteolytic enzymes. From the structural studies of other aspartic proteinase-inhibitor complexes (Bott *et al.*, 1982; James *et al.*, 1982; Cooper *et al.*, 1989; Šali *et al.*, 1989; Bailey *et al.*, 1993), it is known that this flap is involved in the substrate binding. Therefore, the destruction of this flap by proteolysis may reduce substrate binding and result in the loss of enzyme activity. Since the surface loop where Asn188 is located has greater positional freedom than β-strands or β-turns, the carbohydrate moiety attached to Asn188 is not sterically restricted. This lack of steric restraint allows this carbohydrate moiety to block access of proteolytic enzymes, thereby conferring protection from proteolysis.

Recent studies on glycosylated enzymes and glycoproteins have shown that N-linked carbohydrates enhance the thermal stability of glycoproteins (Doan & Fincher, 1992; Kodama *et al.*, 1993; Chen *et al.*, 1994; Terashima *et al.*, 1994; O *et al.*, 1995). It has been proposed that the N-linked carbohydrate chains stabilize the tertiary structure of the glycoproteins (O *et al.*, 1995). One interpretation for this phenomenon is that the carbohydrate chains on the glycoproteins have large mobilities. The mobilities of the carbohydrate chains increase as the temperature rises. It is likely that the vibrational motions of the carbohydrate chains act as heat reservoirs and confer higher thermal stability on the glycoproteins. Denaturation studies by

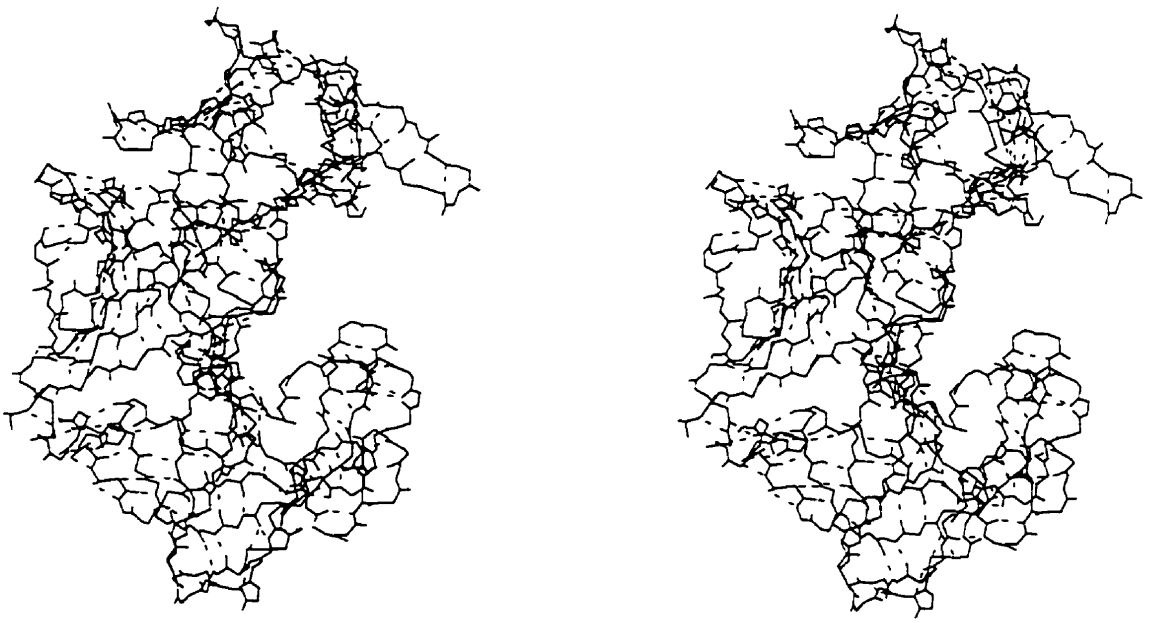
Brown & Yada (1991) show that the activation energies for the denaturation of glycosylated RMP and partially deglycosylated (37.8%) RMP are 458 kJ/mol and 339 kJ/mol, respectively. These studies are consistent with the hypothesis that the carbohydrates act as heat reservoirs in thermally stable glycoproteins. The carbohydrates may also prevent the aggregation of the transition state (partially-unfolded) forms of the glycoproteins during the denaturation process and prolong the presence of the native forms (Brown & Yada, 1991). The above rationale may explain why RMP exhibits very high thermal stability.

Deglycosylation studies by Aikawa *et al.* (1990) have shown that the depletion of the N-linked carbohydrate groups from MPP increases the milk-clotting activity and decreases the proteolytic activity and thermal stability of the enzyme. Recent studies by Harboe (1996) indicated that RMP deglycosylated by endo-β-N-acetylglucosaminidase H had an increase of 30% of the clotting activity. Unfortunately, the thermal stability change was not reported. However, due to the similarity between RMP and MPP, it is likely that deglycosylation of RMP by endo-β-N-acetylglucosaminidase H will also result in a decrease in thermal stability.

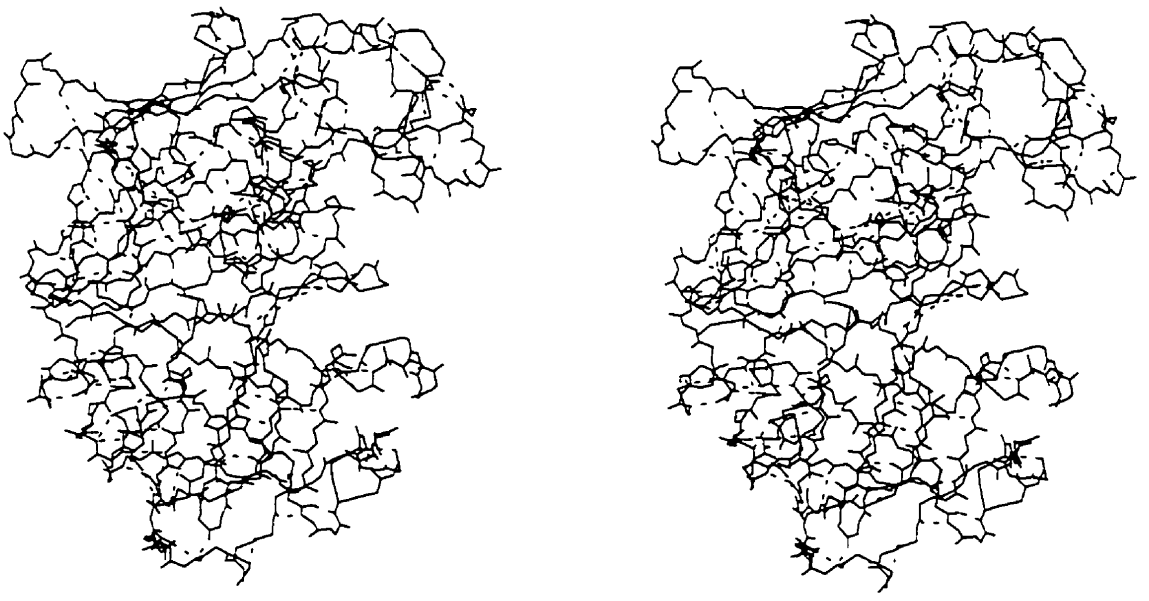
The results of mutagenesis studies by Yamashita *et al*, (1994) indicate that the substitution of Gly186 by various other amino acids in MPP decreases the thermal stability of the enzyme. Glycine does not have torsion angle restraints and can be accommodated where other amino acids cannot fit. Mutation of Gly186 to any other amino acid would probably increase the strain in that region and cause the protein to be less stable, and therefore the thermal stability of the enzyme would be decreased. Gly186 in the RMP molecule is in a "non-allowed" region of Ramachandran Plot. The possibility that the change at position 186 may decrease the mobility of the carbohydrate chain at Asn188 cannot be eliminated. It is expected that similar mutants of RMP would also show a decrease in thermal stability, due to the high structural similarity of RMP and MPP.

The study of the thermal stabilities of proteins is a very important research area and has attracted the attention of many scientists for many years. Different explanations have been proposed for the thermal stabilities of the proteins, including increased hydrophobic core interaction and increased interdomain interface area (Korolev *et al.*, 1995), disulfide bonds (Betz, 1993), and internal hydrogen bonding network (Yu *et al.*, 1995). All aspartic proteinases consist predominantly of β-sheet structures, have a hydrophobic core and a rigid active site. The network of hydrogen bonds between

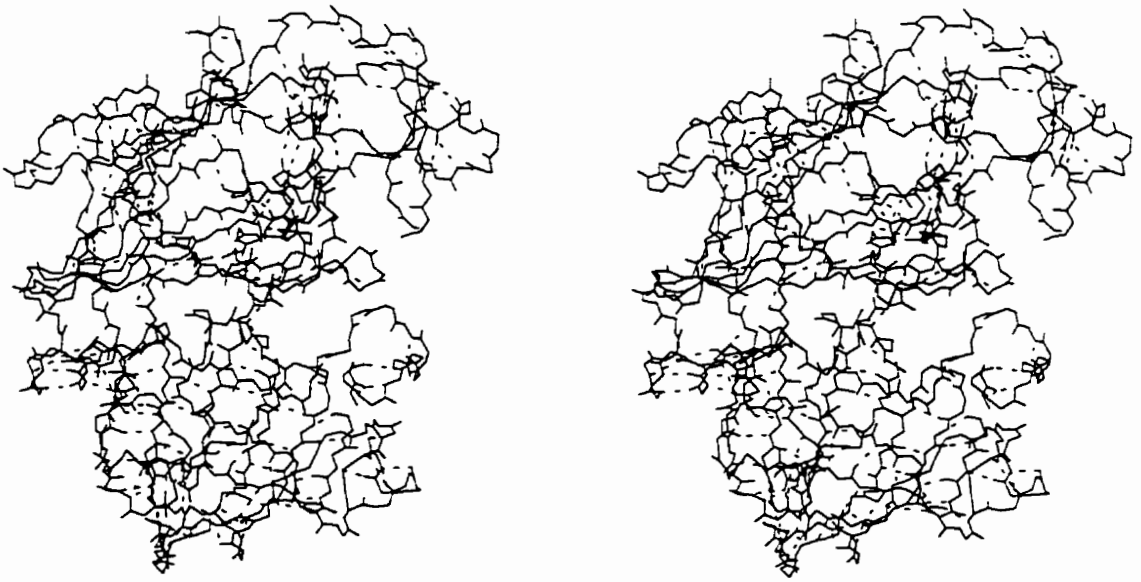
β-strands, the network of hydrogen bonds in the active site, and the hydrophobic interactions contributed by the hydrophobic core might be the reasons for the high thermal stability of the enzymes. To understand better the reason for the thermal stability of RMP, the structures of chymosin B, RMP, MPP and renin were examined. Heat inactivation studies (Hyslop *et al.*, 1979) show that the temperatures necessary to inactivate the enzymes by 90% at pH 5.0 in 15 seconds are 82°C, 74°C and 72°C for RMP, MPP and rennet (chymosin), respectively. The in vitro studies by De Vito *et al.* (1997) showed that renin was rapidly inactivated in rat plasma at 37°C. No significant differences in hydrogen bonding and hydrophobic interaction were found among these four enzymes in the comparison studies. The main chain hydrogen bonds in these four enzymes are shown in Figure 3.16. The deglycosylation studies by Aikawa *et al.* (1990) on MPP also eliminate the possibility that hydrogen bonding networks and hydrophobic interactions supply the enzyme with high thermal stability, since the deglycosylation does not change the tertiary structure of the enzyme. In a recent minireview by Goldman (1995), a network of ionic pairs in proteins was suggested to be the reason for the high thermal stability of the proteins. However, a thorough examination of the RMP structure did not find such a network of ionic pairs in RMP. Therefore, it can be concluded that the high thermal stability of RMP is primarily conferred by the N-linked carbohydrates.



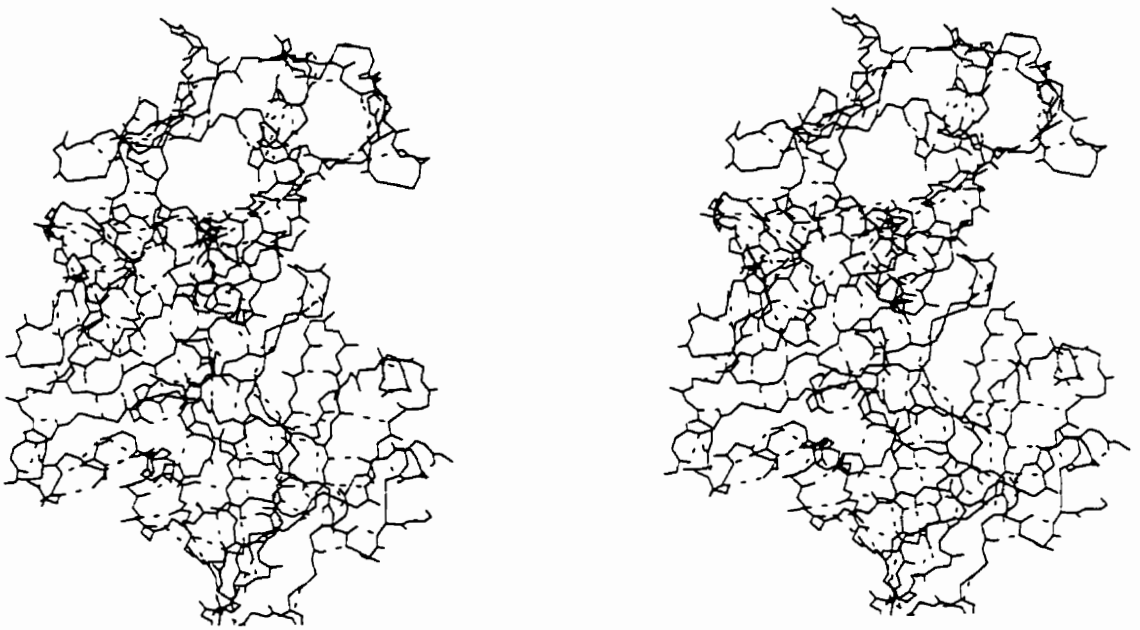
(a) chymosin B



(b) RMP



(c) MPP



(d) renin

Figure 3.16. Main chain hydrogen bonds in (a) chymosin B, (b) RMP, (c) MPP and (d) renin. These diagrams are in stereo.

3.2.7 Structure comparison with other aspartic proteinases

The amino acid sequence alignment of RMP, MPP, chicken pepsin, chymosin B, penicillopepsin, endothiapepsin and yeast proteinase A was done by Baudyš *et al.* (1988). This study shows that RMP has 87% identity with MPP and about 22 to 24% identity with other aspartic proteinases. In the present study the amino acid sequence alignment of RMP with MPP, penicillopepsin, endothiapepsin, chymosin B, human renin and pig cathepsin D was carried out. The sequence alignment result is shown in Figure 3.17. The result is consistent with the result from Baudyš *et al.* (1988) that RMP and MPP are structurally closely-related and are significantly different from other aspartic proteinases. After solving the native RMP structure, the final α-carbon atomic coordinates were submitted to EMBL, Heidelberg, Germany for three-dimensional structural alignments with the DALI program (Holm & Sander, 1993). The alignment results are shown in Table 3.10. From this table it can be seen that RMP has a r.m.s. deviation of 0.6 Å from MPP and r.m.s. deviation of more than 1.7 Å from the other aspartic proteinases for 215 to 226 equivalent amino acid residues. This indicates that RMP and MPP diverged from other aspartic proteinases at an early stage of evolution and form a sub-family within the aspartic proteinases. This may explain why

```
RMP_sequen      AAADGSVDTPGYYDFDLEEYAIPVSIGTPGQDFLLLFDTGSSDTWVPHKGCTKSE-GCVG
MPP_sequen      AEGDGSVDTPGLYDFDLEEYAIPVSIGTPGQDFYLLFDTGSSDTWVPHKGCDNSE-GCVG
Penicillop      --AASGVATNTPTAN-DEEYITPVTIGG--TTLNLNFDTGSADLWVFSTELPAS--QQSG
Endothiape      ---STGSATTTPIDSLDDAYITPVQIGTPAQTLNLDFDTGSSDLWVFSSETTAS--EVDG
Chymosin_b      ----GEVASVPLTNYLDSQYFGKIYLGTPPQEFTVLFDTGSSDFWVPSIYCKSN--ACKN
Human_reni      LTLGNTTSSVILTNYMDTQYYGEIGIGTPPQTFKVVFDTGSSNVWVPSSKCSRLYTACVY
Pig_cathep      -----GPIPEVLKNYMDAQYYGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCKLLDIACWI
                                    *   . .*       . *****.  **

RMP_sequen      SRFFDPS-ASSTFKATNYNLNITYGTG-GANGLYF----------EDSIAIGDITVTKQI
MPP_sequen      KRFFDPS-SSSTFKETDYNLNITYGTG-GANGIYF----------RDSITVGGATVKQQT
Penicillop      HSVYNP--SATGKELSGYTWSISYGDGSSASGNVF----------TDSVTVGGVTAHGQA
Endothiape      QTIYTPSKSTTAKLLSGATWSISYGDGSSSSGDVY----------TDTVSVGGLTVTGQA
Chymosin_b      HQRFDPR-KSSTFQNLGKPLSIHYGTG-SMQGILG----------YDTVTVSNIVDIQQT
Human_reni      HKLFDAS-DSSSYKHNGTELTLRYSTG-TVSGFLS----------QDIITVGGITVT-QM
Pig_cathep      HHKYNSG-KSSTYVKNGTTFAIHYGSG-SLSGYLSSQDTVSVPCNSALSGVGGIKVERQT
                   .     ..           . *  *    *                    .       *

RMP_sequen      LAYVDNVRGPTAEQSPNADIFLDGLFGAAYPDNTAMEAEYGSTYNTVHVNLYKQGLISSP
MPP_sequen      LAYVDNVSGPTAEQSPDSELFLDGMFGAAYPDNTAMEAEYGDTYNTVHVNLYKQGLISSP
Penicillop      VQAAQQI--SAQ---FQQDTNNDGLLGLAFSSINTVQPQS---QTTFFDTVKSS--LAQP
Endothiape      VESAKKV--SSS---FTEDSTIDGLLGLAFSTLNTVSPTQ---QKTFFDNAKAS--LDSP
Chymosin_b      VGLSTQE--PGD---VFTYAEFDGILGMAYPSLASEYSIP------VFDNMMNRHLVAQD
Human_reni      FGEVTEM--PAL---PFMLAEFDGVVGMGFIEQAIGRVTP------IFDNIISQGVLKED
Pig_cathep      FGEATKQ--PGL---TFIAAKFDGILGMAYPRISVNNVVP------VFDNLMQQKLVDKD
                                 **. *  .                    .       .

RMP_sequen      LFSVYMNTN--SGT---GEVVFGGVNNTLLGGDIAYTDVMSRYGGYYFWDAPVTGITVDG
MPP_sequen      VFSVYMNTN--DGG---GQVVFGGANNTLLGGDIQYTDVLKSRGGYFFWDAPVTGVKIDG
Penicillop      LFAVALKHQ---QP---GVYDFGFIDSSKYTGSLTYTGVDNSQG---FWSFNVDSYTAGS
Endothiape      VFTADLGYH---AP---GTYNFGFIDTTAYTGSITYTAVSTKQG---FWEWTSTGYAVGS
Chymosin_b      LFSVYMDRN--GQE---SMLTLGAIDPSYYTGSLHWVPVTVQQ----YWQFTVDSVTISG
Human_reni      VFSFYYNRDSENSQSLGGQIVLGGSDPQHYEGNFHYINLIKTG----VWQIQMKGVSVGS
Pig_cathep      IFSFYLNRD-PGAQ-PGGELMLGGIDSKYYKGSLDYHNVTRKA----YWQIHMNQVAVGS
                .*.                       *        *      .         *
```

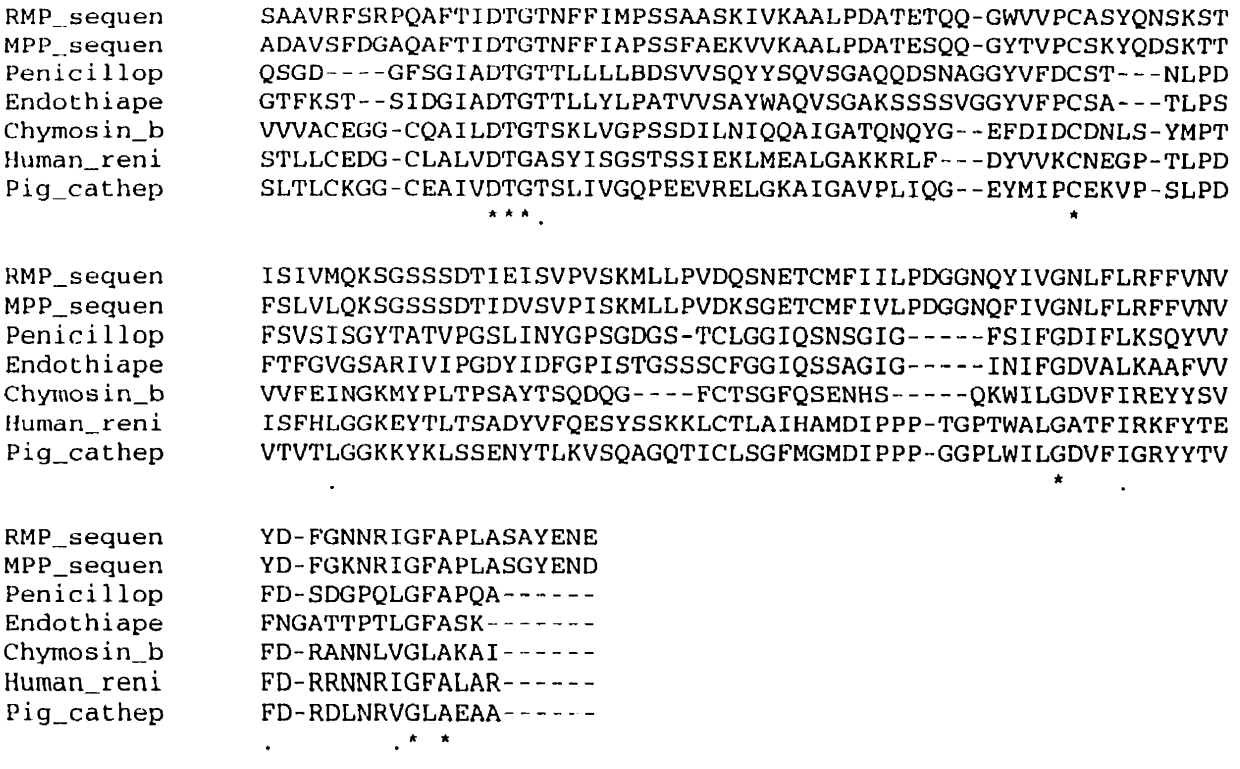


Figure 3.17. Sequence alignment of RMP with MPP, penicillopepsin, endothiapepsin, chymosin b, human renin 3A and pig cathepsin D. The amino acid sequences were obtained from the SWISS-PROT protein sequence data bank. Identical residues among the aligned sequences are highlighted with an asterisk (*), and conserved residues with a dot (·).

molecular replacement search with aspartic proteinases other than MPP did not give a solution for RMP and MPP was only able to be solved by multiple isomorphous replacement (MIR) method. RMP and MPP maintain the overall structural similarity of aspartic proteinases: a large substrate-binding cleft, two catalytically essential aspartate residues in the middle of the cleft, mobility of the two domains, relative rigidity of the active site due to hydrogen bonding, and a highly mobile flap region over the substrate-binding cleft. However, due to the differences in the residues in the substrate-binding site, the aspartic proteinases may show diverse substrate specifities (Yang *et al.*, 1997).

Table 3.10. The r.m.s. deviations between RMP and other aspartic proteinases from the three-dimensional structural alignments

Aspartic proteinases source	R.m.s. deviation from RMP (Å)
MPP	0.6
human pepsin	1.7
rhizopuspepsin	1.8
chymosin B	1.8
human renin	1.8
penicillopepsin	2.0
endothiapepsin	2.2

3.3 Conclusion

The crystal structure of the aspartic proteinase from *Rhizomucor miehei* (RMP, EC 3.4.23.23) has been refined at 2.15 Å resolution to a crystallographic *R*-factor of 21.5% and a *R*-free of 28.1%. The root-mean-square (r.m.s.) error for the atomic coordinates estimated from a Luzzati plot is 0.2 Å. The r.m.s. deviations for the bond distances and bond angles from the constraints in X-PLOR are 0.01 Å and 1.7°, respectively. RMP contains two domains, which consist predominantly of β-sheets. A large substrate binding cleft is clearly visible between the two domains, and the two catalytic residues Asp38 and Asp237 are located in the middle of the cleft with a water molecule bridging the carboxyl groups of Asp38 and Asp237. Due to crystal packing, the C-terminal domain is more mobile than the N-terminal domain. Most of the aspartic proteinases (except renin) reach their maximum activity at acidic pH. From this study it is proposed that the optimum pH of each aspartic proteinase is determined by the electrostatic potential at the active site, which, in turn, is determined primarily by the positions and orientations of all the residues near the active site. RMP is the most glycosylated among the aspartic proteinases. The carbohydrate moieties are linked to Asn79 and Asn188, respectively. Asn79 is in the middle of a β-strand and Asn188 is on a

surface loop in contrast to the previous hypothesis proposed by Brown and Yada (1991) that they are both on the surface β-turns. RMP has a very high thermal stability. The high thermal stability is probably due to the high level of glycosylation. It is proposed that the highly flexible carbohydrates act as heat reservoirs to stabilize the conformation of RMP and therefore provide the enzyme with high thermal stability. Three-dimensional structural and sequence alignments of RMP with other aspartic proteinases show that RMP is most structurally homologous to MPP, and differs from other fungal enzymes as much as it does from the mammalian enzymes. This suggests that RMP and MPP diverged from the main stream of aspartic proteinases at an early stage of evolution. The present study adds a second member to this subfamily of aspartic proteinases.

Chapter 4. Crystal Structure of RMP-pepstatin A complex at 2.7 Å

During the past twenty years, aspartic proteinases have been extensively studied due to their diverse biological functions. All members of the aspartic proteinase family are closely related structurally. They all contain two domains (homodimers for retroviral aspartic proteinases) with the active site at the interface of these two domains. In order to determine the catalytic mechanism of this group of enzymes, the structures of many aspartic proteinases complexed with natural or synthetic inhibitors have been studied (James & Sielecki, 1985; Suguna *et al.*, 1987; Foundling *et al.*, 1987a,b; Cooper *et al.*, 1989; Abad-Zapatero *et al.*, 1991; Hoover *et al.*, 1991; Pitts *et al.*, 1995). These results indicate that a large substrate-binding cleft, which can accommodate up to eight amino acid residues, is situated between the two domains, and the inhibitors fit into this substrate-binding cleft in extended conformations. The residue at P1 (or P1-P1') of the inhibitors interacts with the two catalytic aspartates by hydrogen bonds, and this type of interaction is believed to resemble the tetrahedral intermediate in proteolytic cleavage of the peptide bond (Veerapandian *et al.*, 1992).

The studies of aspartic proteinases complexed with various inhibitors have become a very active area with the discovery that HIV proteinases are aspartic proteinases. Many competitive inhibitors of HIV proteinases have been designed. Three proteinase inhibitors, namely, indinavir, ritonavir and saquinavir, with potent anti-HIV activity and favorable pharmacological properties have been approved by the United States Food and Drug Administration (FDA) for use in humans for the treatment of AIDS. These inhibitors can greatly reduce the number of new, infectious copies of HIV made inside cells. If these inhibitors succeed in making most new HIV viruses defective, the HIV infection would not spread inside the body as quickly as it does now. However, due to mutation, HIV viruses show resistance to these inhibitors (Otto *et al.*, 1993; Ho *et al.*, 1994; Kaplan *et al.*, 1994; Condra *et al.*, 1995; Markowitz *et al.*, 1995). Combinations of HIV proteinase inhibitors and combinations of the HIV proteinase inhibitors with the reverse transcriptase inhibitors have been studied as potential treatments for AIDS. However, these studies are still at the beginning stage and it is too early to say which combinations will work best. Combinations of inhibitors may increase known side effects or cause new ones (Markowitz, 1996). Very recently, Mr. Bill Clinton, the President of the United States of America, called for development of an AIDS vaccine (Nature, 1997, Vol. 387, p323). HIV proteinases are still

the most promising targets and a potent inhibitor, which resists HIV mutations, may be a breakthrough in the the search for a cure for AIDS.

In Chapter 3, it was mentioned that RMP and MPP diverged from other aspartic proteinases at an early stage of evolution and form a sub-family of the aspartic proteinases. Therefore, the binding of inhibitors to RMP and MPP might be different from other aspartic proteinases. As no structures of RMP or MPP complexes with inhibitors have been published, the structural study of RMP complexed with pepstatin A was carried out. The present study is the first structure of an enzyme-inhibitor complex in this sub-family of aspartic proteinases and this research will also improve our knowledge of the aspartic proteinases.

4.1 Structure solution and refinement

4.1.1 Crystallization

The crystallization of RMP-pepstatin A complex was a tedious process. The complex was crystallized by a method which involves the combination of repeated soaking and seeding. The inhibitor, pepstatin A, was purchased from Calbiochem-Behring Corp., USA. Because of the low solubility of pepstatin A in water, only 1:1 molar ratio of pepstatin A to RMP was used in the crystallization. Crystals of the native enzyme were grown by the hanging drop vapour diffusion method with 25 mg/mL RMP solution (20 mM sodium citrate buffer, pH 3.6, 8% w/v PEG 8000) suspended over 26% w/v PEG 8000 reservoir solution (20 mM sodium citrate buffer, pH 3.6) at 21°C. These thin needle-shaped native enzyme crystals were used as the initial seeds for the repeated combination of soaking and re-seeding (Thaller *et al.*, 1981), which was carried out by the sandwich drop method. Aliquots of 10 μL precipitation solution (20 mM citrate buffer, pH 3.6, 26% w/v PEG 8000) were mixed with 20 μL aliquots containing protein (25 mg/mL RMP, 0.45 mg/mL pepstatin A, 20 mM citrate buffer, pH 3.6). The seed crystals then were put into the mixed protein drops to equilibrate over the precipitation solution. After about one to

two months, the crystals, with pepstatin A molecules soaked into the active sites of RMP molecules, had grown a little larger. These larger crystals were used as seeds for another cycle of soaking and re-seeding. The crystals used for data collection were obtained after five cycles of this combination of soaking and seeding. The size of the crystal used for the synchrotron data collection was approximately 0.40 × 0.25 × 0.05 mm. The crystals of the RMP-pepstatin A complex grow in the same crystal system as does the native enzyme. The space group is $P2_12_12_1$, and the unit cell dimensions are: a = 41.52 Å, b = 50.82 Å, c = 172.71 Å, $\alpha = \beta = \gamma = 90°$. This unit cell is slightly smaller than the unit cell of the native enzyme (Yang *et al.*, 1997). There is one RMP-pepstatin A complex molecule per asymmetric unit.

4.1.2 Data collection

Two crystals of the RMP-pepstatin A complex were mounted in glass capillary tubes of 1.0 mm in diameter and were taken to the Brookhaven National Laboratory, New York, USA, for synchrotron data collection on beamline X12C. The data set was collected with only one crystal. The other crystal was damaged during the transportation. The X-ray data of the complex were collected with a Brandeis CCD detector to a maximum

resolution of 2.7 Å. The wavelength of the incident radiation was 1.15 Å. The data were processed using DENZO and SCALEPACK (Otwinowski, 1993). A total of 43156 reflections measured to 2.7 Å were merged to yield 9360 unique reflections. The merging R_{sym} is 0.092 for symmetry-equivalent reflections based on intensities. The completeness of the data set is 84.9% up to 2.7 Å. The R_{sym} and data completeness for each resolution shell are shown in Table 4.1.

4.1.3 Model building and refinement

The coordinates of the native RMP enzyme were used as the starting model. In the first cycle, the model was refined with all of the reflections within the resolution range 8 to 2.7 Å, using the X-PLOR package (Brünger, 1992b). The following protocols were used in the refinement: (1) check step (to obtain the weights for diffraction data); (2) rigid body refinement; (3) preparation stage (to relieve strains or bad contacts of the initial coordinates); (4) slow cooling refinement (simulated-annealing refinement) and (5) individual B-factor refinement. The rigid body refinement was first applied the entire RMP molecule and then allowing relative movement of the two domains of the RMP molecule. After the rigid body refinement, the R-value was 28.2%. After the

Table 4.1. R_{sym} and completeness for each resolution shell (I > 0)

Resolution shell (Å)	R_{sym}*	Completeness (%)
50.00 to 5.60	0.235	99.1
5.60 to 4.45	0.039	100
4.45 to 3.88	0.043	99.9
3.88 to 3.53	0.057	99.7
3.53 to 3.28	0.072	99.9
3.28 to 3.08	0.094	96.1
3.08 to 2.93	0.120	80.0
2.93 to 2.80	0.154	64.5
2.80 to 2.70	0.159	47.9
Overall	0.092	84.9

* $R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma \langle I \rangle$

B-factor refinement, the *R*-value dropped to 25.0%. In the slow cooling refinement, the system was slowly cooled from 2500 K to 300 K in the steps of 50 K with energy minimization in each step. SIGMAA weighted $2F_o$-F_c and ΔF (F_o-F_c) maps (Read, 1986) were calculated after the refinement. The maps were examined using the TURBO-FRODO program (Roussel *et al.*, 1990) and

the position of the pepstatin A molecule was located from the ΔF map. The electron density for the pepstatin A molecule was observed at 2σ contour level in the ΔF map. The model of the complex was then subjected to six cycles of simulated-annealing refinement and sixteen cycles of positional refinement with all reflections in the resolution range 8 to 2.7 Å. Powell minimization was used in the preparation stage, after simulated-annealing refinement in the slow cooling stage and in the positional refinement stage. The refinements were interspersed with manual model building and adjustments using the program TURBO-FRODO. The R-free (test set comprised about 5% of the observed data) dropped from 34.2% of the second cycle of refinement to 28.0% of the second last cycle of refinement. In the final cycle of refinement, all reflections were used and the R-value was 19.3%. No solvent molecules were added to the model. The final results of the refinement are listed in Table 4.2.

Table 4.2. X-PLOR refinement results

Number of protein non-H atoms	2665
Number of carbohydrate non-H atoms	39
Number of inhibitor non-H atoms	48
Resolution range (Å)	8.0 - 2.7
Number of unique refs. used in refinement	8599
Conventional *R*-value (%)	19.3
R-free (%)	28.0
R.m.s.d. for bond distances (Å)	0.01
R.m.s.d. for bond angles (degree)	1.7
R.m.s.d. for dihedral angles (degree)	27.2
R.m.s.d. for impropers (degree)	1.8
Average B-factor for the enzyme ($Å^2$)	32.3
Average B-factor for N-terminal domain ($Å^2$)	25.1
Average B-factor for C-terminal domain ($Å^2$)	39.8
Average B-factor for carbohydrate ($Å^2$)	70.7
Average B-factor for the inhibitor ($Å^2$)	76.1
Average B-factor for the active site surface flap ($Å^2$)	51.7
Average B-factor for the flexible sub-domain ($Å^2$)	45.3
Average B-factor for the residues in the C-terminal domain except those of the flexible sub-domain ($Å^2$)	28.6

4.2 Results and discussion

4.2.1 Quality of the final model and overall folding

The structure of RMP complexed with pepstatin A has been refined to a crystallographic *R*-value of 19.3% and a *R*-free of 28.0% at 2.7 Å resolution. The error for the atomic coordinates estimated from a Luzzati plot (Luzzati, 1952) is 0.3 Å. The final results of the refinement are summarized in Table 4.2. Three N-terminal residues and the side chains of fifteen residues were omitted in the refinement due to insufficient indication of their positions in the final electron density map. These residues are listed in Table 4.3. An electron density map ($2F_o$-F_c) at the inhibitor pepstatin A site is shown in Figure 4.1. The final model was also examined for the main chain torsion angles (ϕ, ψ) by the program PROCHECK (Laskowski *et al.*, 1993). A Ramachandran plot (Ramachandran & Sasisekharan, 1968) produced by the program PROCHECK is shown in Figure 4.2. Except for the glycine and the proline residues, all of the amino acid residues are in the allowed regions.

A ribbon representation of the RMP molecule complexed with pepstatin A is given in Figure 4.3. As predicted in Chapter 3, pepstatin A fits into the substrate-binding cleft between the two domains of RMP in an

Table 4.3. Omitted regions in RMP-pepstatin A complex

Ala1 to Ala3	All atoms
Asp4	Side-chain atoms
Lys49	Side-chain atoms
Lys53	Side-chain atoms
Thr84	Side-chain atoms
Glu120	Side-chain atoms
Asn124	Side-chain atoms
Arg203	Side-chain atoms
Ser223	Side-chain atoms
Ser247	Side-chain atoms
Ser251	Side-chain atoms
Gln265	Side-chain atoms
Ser274	Side-chain atoms
Lys279	Side-chain atoms
Ser291	Side-chain atoms
Val309	Side-chain atoms

extended conformation. The average B-factor for the RMP-pepstatin A complex molecule (32.3 $Å^2$) is 6.8 $Å^2$ lower than that for the native enzyme molecule (39.1 $Å^2$); this suggests that the complex is slightly more rigid than the native molecule. The average B-factor for the flexible C-terminal

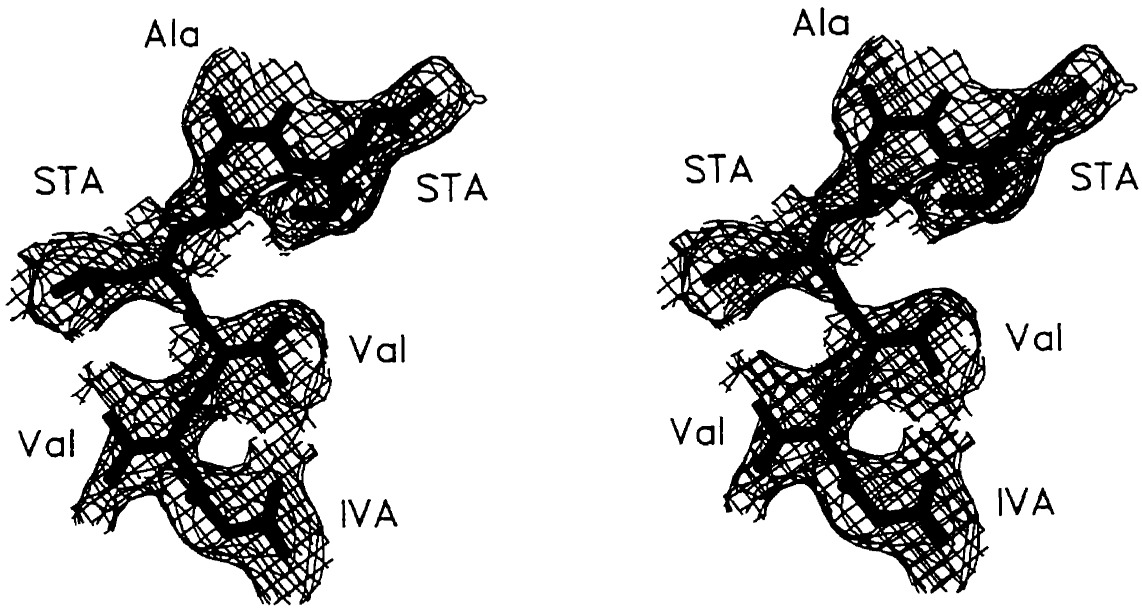


Figure 4.1. Electron density map ($2F_o$-F_c) at the pepstatin A site. This figure is prepared by SETOR (Evans, 1993).

sub-domain in the complex is about 9.5 Å^2 lower than in the native enzyme structure. This indicates that the flexibility of this sub-domain decreases due to the binding of the pepstatin A molecule. However, the average B-factor for this flexible sub-domain is still much higher than the average B-factor for the N-terminal domain in the RMP-pepstatin A complex structure. This indicates that this sub-domain is still flexible compared to the N-terminal domain. The reason for the flexibility of this sub-domain is probably that the complex was

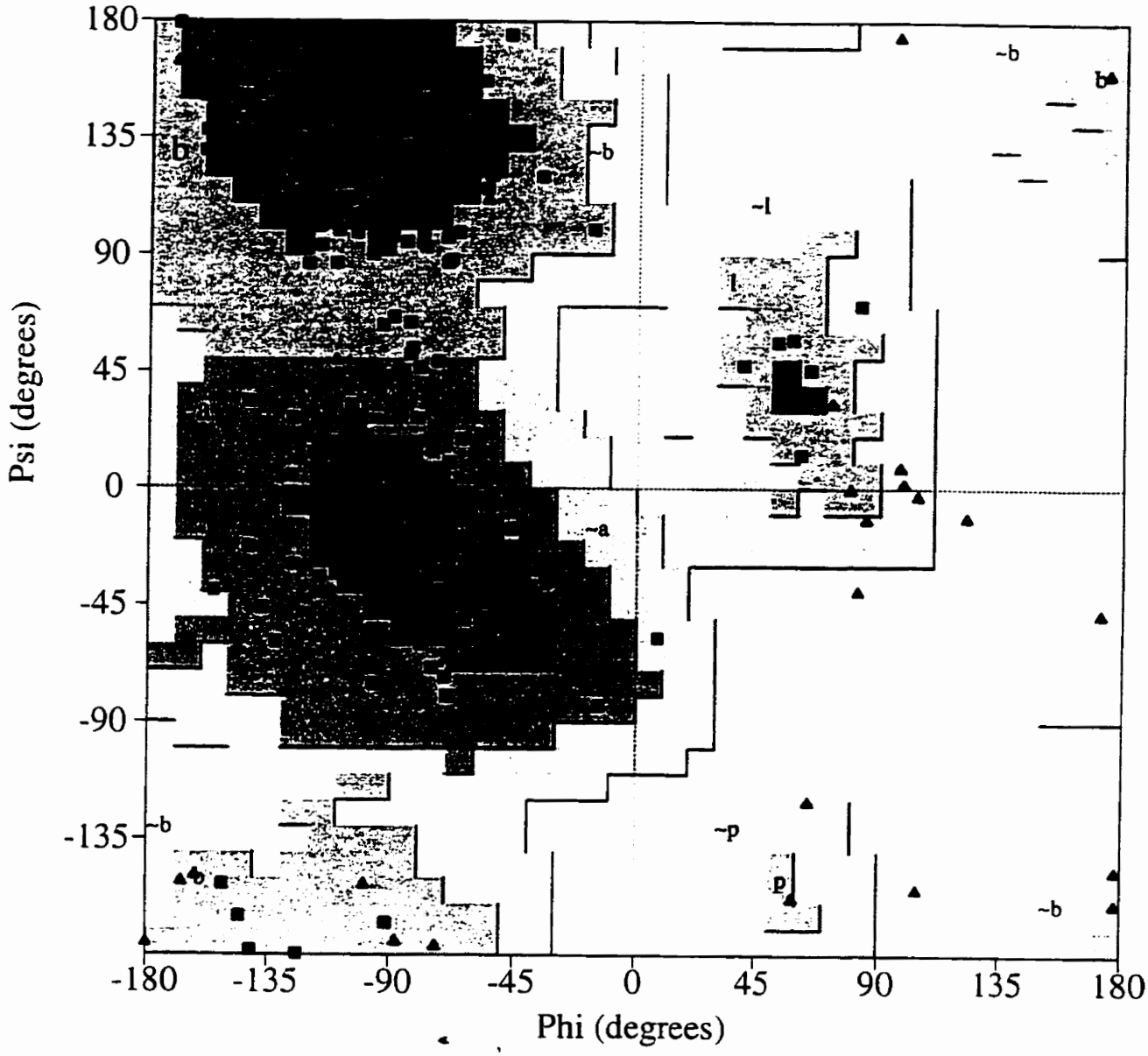


Figure 4.2. A Ramachandran plot of RMP-pepstatin A complex, showing the main chain torsion angles (ϕ, ψ). The glycines are shown as triangles.

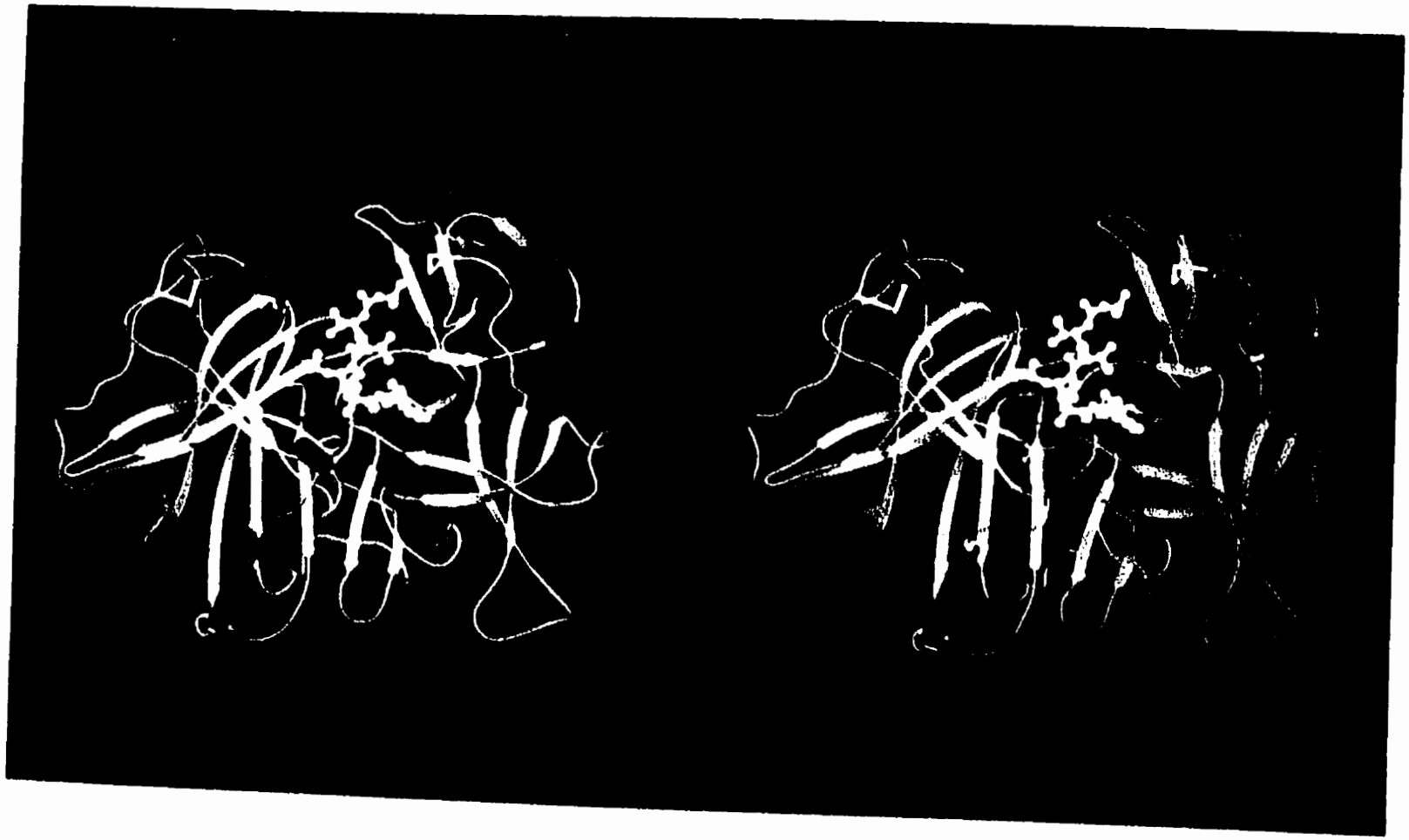


Figure 4.3. A stereo ribbon representation of the RMP-pepstatin A complex.

crystallized by soaking native crystals with pepstatin A. The crystal structure of the complex is the same as the native enzyme and there is no change in the interactions between one RMP molecule and its symmetry-related molecules, which were presumed to be responsible for differences in mobility of the N and C-terminal domains in RMP.

4.2.2 Comparison with the native enzyme

In order to examine any conformational change of RMP due to the binding of pepstatin A, the coordinates of the RMP-pepstatin A complex structure were superimposed with the coordinates of the native enzyme (Figure 4.4). The two domains of RMP did not show any movement in the complex structure with respect to the native enzyme. No movement of the sub-domain was observed either. As only the initial and final stages of the inhibitor binding were observed, it is not known whether there were any domain movements during the binding process. The average B-factor for the flexible sub-domain decreased from 54.8 $Å^2$ in the native enzyme to 45.3 $Å^2$ in the complex. This indicates that the sub-domain becomes less flexible due to the binding of pepstatin A. The binding of a pepstatin A molecule does not cause any large distortion of the active site of RMP. Only the active-site

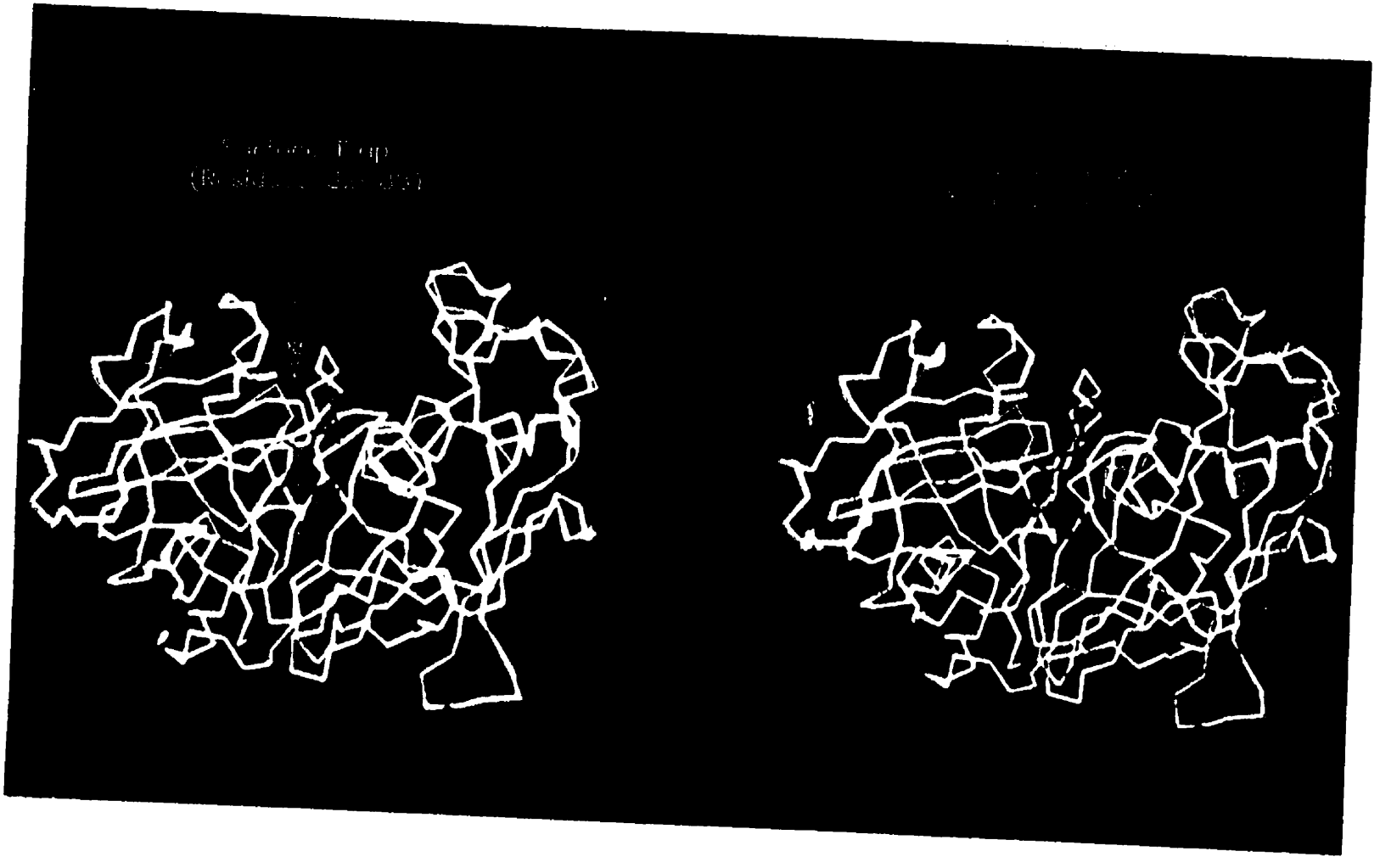


Figure 4.4. Superposition of the complex structure with the native enzyme structure (stereo view). The native enzyme, the complexed enzymes, and the inhibitor pepstatin A are shown in blue, yellow and red, respectively.

surface flap (residues 82 to 88) and a surface loop (residues 323 to 328) underwent conformational changes. The conformational change of the active-site surface flap is shown in Figure 4.5. The flap moved towards the inhibitor and formed two hydrogen bonds with the inhibitor. The flap covers up positions P2-P1'. Its residues are involved in the formation of subsites S2, S1 and S2'. The average B-factor of the flap (51.7 $Å^2$) in the complex structure is much lower than in the native structure (66.6 $Å^2$), and this suggests the flap becomes more rigid in the complex. The conformational change of the surface loop is shown in Figure 4.6. This surface loop is opposite to the surface flap at the active site. It shifts about 1 to 1.5 Å towards the N-terminus of the pepstatin A molecule. Gly325 makes good van der Waals contacts with the P2 valine residue of pepstatin A. The average B-factors for this surface loop in the native and complex structures (68.1 $Å^2$ and 64.8 $Å^2$, respectively) are very similar, which indicates that the loop is still flexible in the structure of the complex. The flexibility of this loop allows the access of different substrates to the substrate-binding site and the release of proteolytic products from the enzyme.

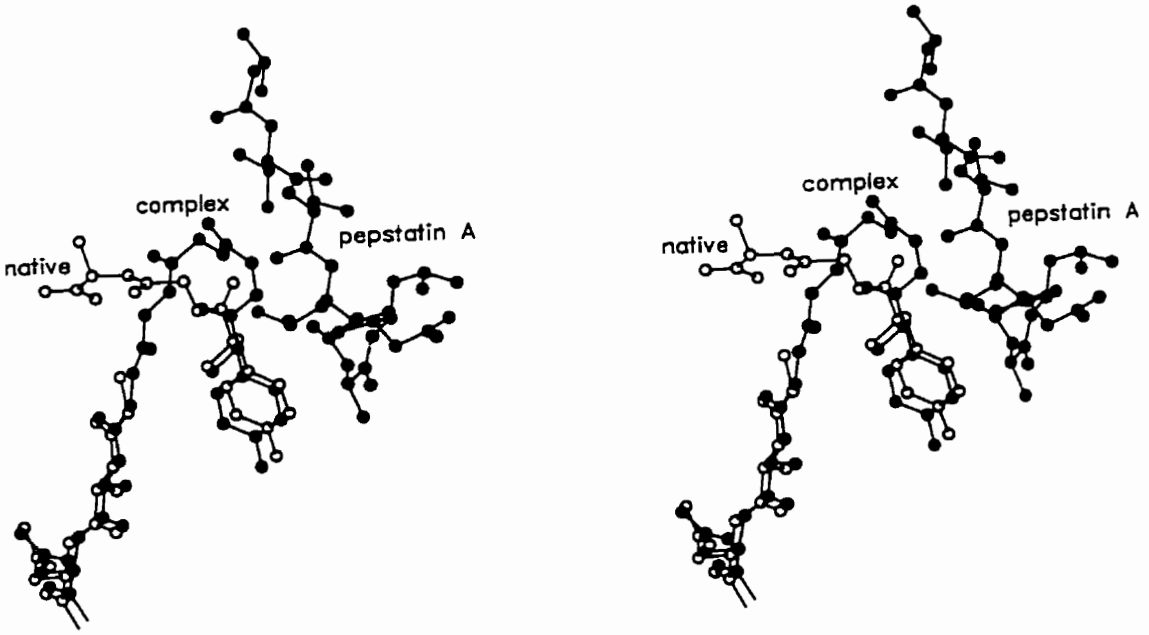


Figure 4.5. The conformational change of the active-site surface flap. Open spheres represent the native conformation. The electron density corresponding to Gly85 in the native enzyme structure was not observed in the native map, but it is present in the map of the complex.

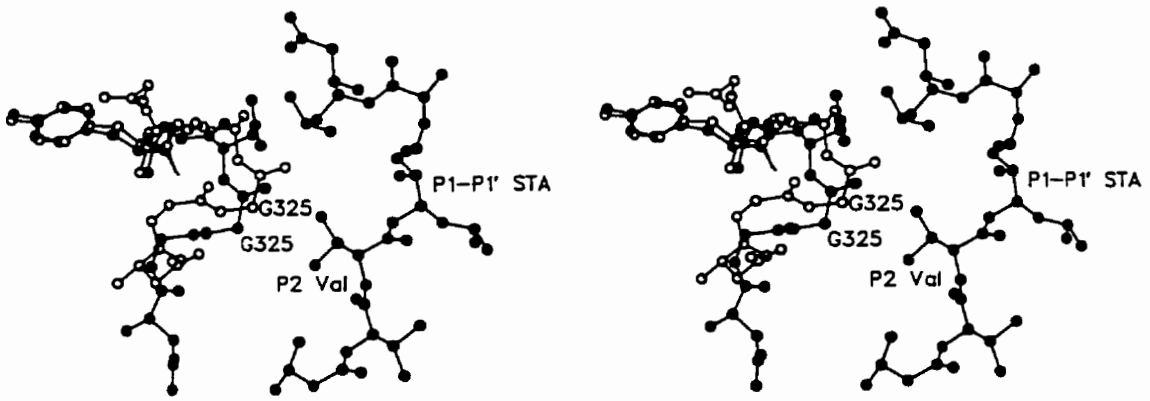


Figure 4.6. The conformational change of the surface loop (residues 323-328). Open spheres represent the native conformation.

4.2.3 Comparison with the results from molecular modeling.

With the development of computer science and increased knowledge of the three-dimensional structures of proteins, molecular modeling has become a very useful tool in structure-function relationship studies. However, the molecular modeling predictions should be tested by X-ray structural studies. In Chapter 3, the pepstatin A molecule was docked into the RMP molecule by molecular modeling. Although the model was subjected to energy minimization, it is not known if the inhibitor pepstatin A conformation was correct. Solution of the RMP-complex structure makes it possible to evaluate the molecular modeling results. In Figure 4.7, the superposition of the molecular structure of the RMP-pepstatin A complex with the molecular model is shown. Molecular modeling did not completely give the correct conformation of pepstatin A. The pepstatin A molecule in the model is closer to the correct conformation at the P4 to P1' region than the P2' to P4' region. The central region (P2-P1') in the pepstatin A molecule is the closest to the correct conformation. This suggests that energy minimization did not reach the global minimum, but was trapped at a local energy minimum state. Also, the active-site surface flap did not move closer to the inhibitor during the docking and energy minimization. Therefore, although molecular modeling

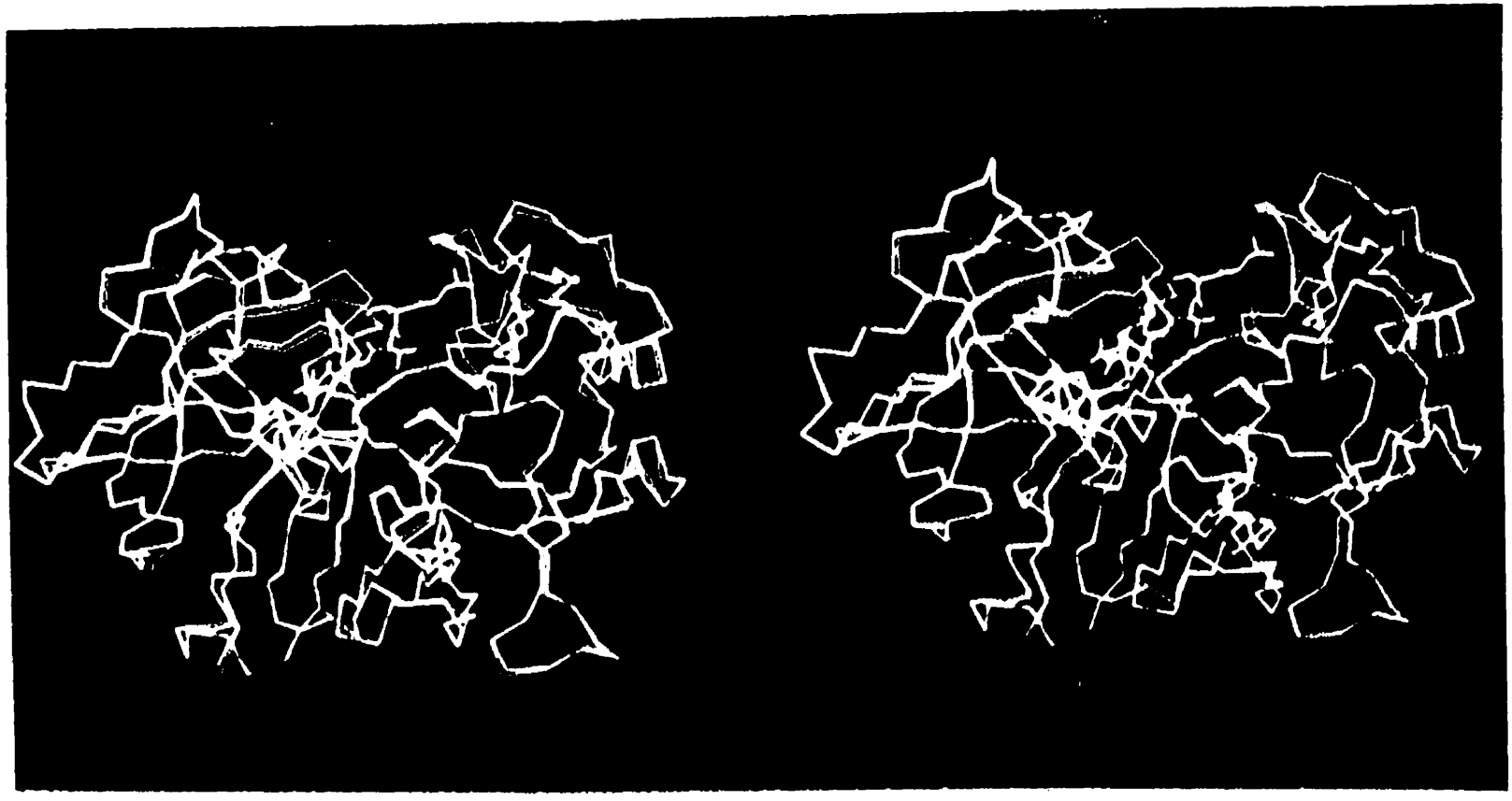


Figure 4.7. The molecular structure of the RMP-pepstatin A complex is superimposed with the molecular modeling model of the complex. The RMP molecule in the model and complex structure are shown in yellow and blue, respectively. The pepstatin A molecule in the model and the complex structure are shown in white and red, respectively.

can provide some useful information it is apparent that X-ray crystallographic studies should be carried out whenever possible.

4.2.4 Hydrogen bonds

The inhibitor pepstatin A makes 10 hydrogen bonds with the enzyme. A schematic diagram of these hydrogen bonds is shown in Figure 4.8. The hydroxyl group at the C3 position of statine P1-P1' replaces the conserved water molecule at the enzyme active site and forms hydrogen bonds with both of the catalytic aspartate residues Asp38 and Asp237 (Figure 4.9). This conformation mimics the expected transition state of the enzyme-substrate reaction. These observations fully support the catalytic mechanism of aspartic proteinases (Hoover *et al.*, 1991; Pearl *et al.*, 1987; Suguna *et al.*, 1987; Veerapandian *et al.*, 1992) which was discussed in Chapter 3. C3 of the statine residue is in the S-configuration. This type of configuration has also been observed in other aspartic proteinase-statine type inhibitor complexes (Cooper *et al.*, 1989; Šali *et al.*, 1989; Bailey *et al.*, 1993). Rich *et al.* (1980) suggested that the 3S enantiomers of the statine residue at P1-P1' position are much more potent than the 3R enantiomers for pepstatin-type inhibitors.

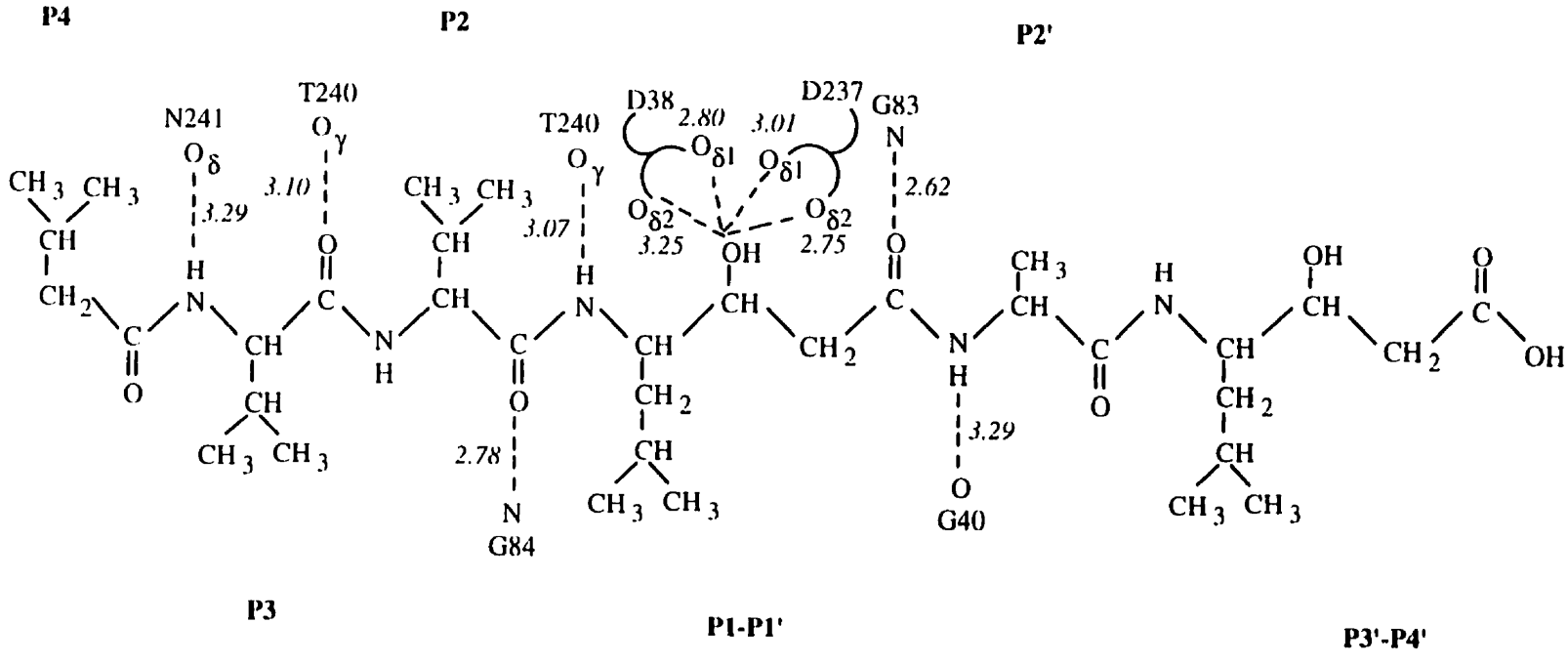


Figure 4.8. A schematic diagram of the hydrogen bonds formed between RMP and pepstatin A.

The active-site flap moves towards the pepstatin A molecule and forms two strong hydrogen bonds with the inhibitor, Gly83 with the P1' CO group and Gly84 with P2 CO group. The formation of these two hydrogen bonds increases the rigidity of this flexible surface flap. The side chain of Thr240 rotates about 90° along the C_α-C_β bond axis to form two hydrogen bonds with the P1 NH and P3 CO groups, respectively. Asn241 forms a hydrogen bond with the P3 NH group, using its O_δ atom, and Gly40 forms a hydrogen bond with the P2' NH group.

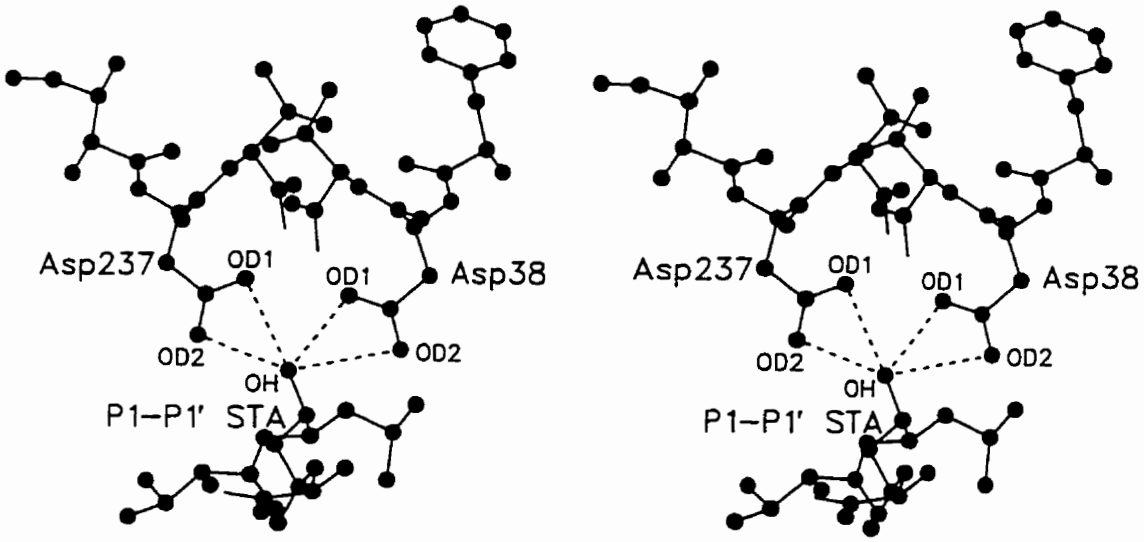


Figure 4.9. The hydrogen bonds formed between pepstatin A and the two catalytically essential aspartate residues.

4.2.5 The γ-turn in pepstatin A

When the CO group of a residue, *i*, forms a hydrogen bond with the NH group of another residue, *i + 2*, a γ-turn is formed. There are two types of γ-turns, classic and inverse (Milner-White *et al.*, 1988). The γ-turn present in pepstatin A is shown in Figure 4.10. The CO group of statine residue at P1-P1' forms a hydrogen bond with the NH group of the statine residues at P3'-P4'. The hydrogen atom in the NH group was placed from geometry of the N atom with the N-H bond distance at 0.97 Å. The O...H and O...N distances are 2.20 Å and 2.86 Å, respectively. The ∠O...H-N angle is 124°. This is an inverse γ-turn. The reason for the formation of this γ-turn in pepstatin A is that the leucyl side of the P3'-P4' statine residue binds back into the S1' subsite because the statine residue at P1-P1' does not have a side chain to occupy subsite S1'. Due to the formation of the γ-turn, the extended conformation of pepstatin A is disrupted at the P3' position. The formation of a γ-turn in the pepstatin A molecule is also observed in other aspartic proteinase-pepstatin A complexes (Suguna *et al.*, 1992; Bailey *et al.*, 1993).

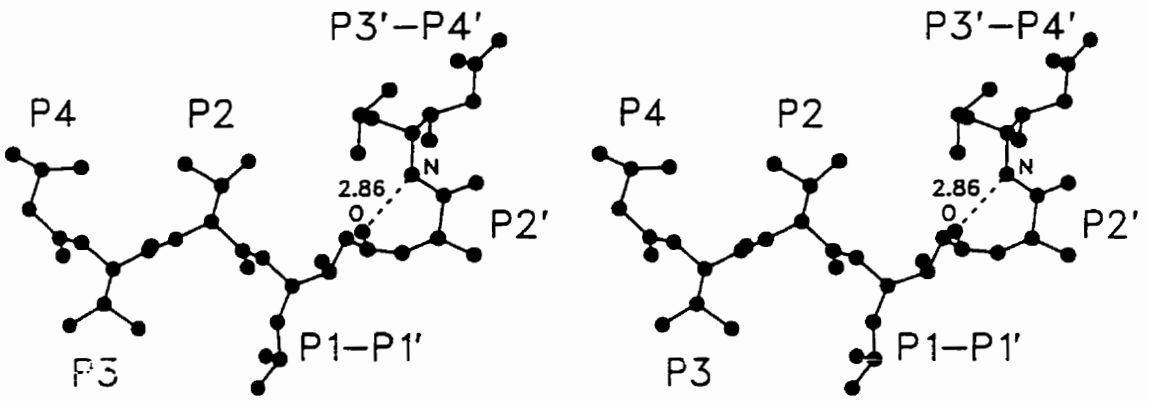


Figure 4.10. The γ-turn in RMP.

4.2.6 Binding subsites

Pepstatin A fits into the substrate-binding cleft between the two domains of RMP mainly in an extended conformation upon binding to the enzyme. All of the substrate-binding cleft residues of the inhibitor complex are shown in Figure 4.11. The subsites' residues, which make van der Waals contacts with the inhibitor ($d \leq 4.0$ Å), are listed in Table 4.4. As mentioned in the previous section, the leucine-like side chain of the statine residue at

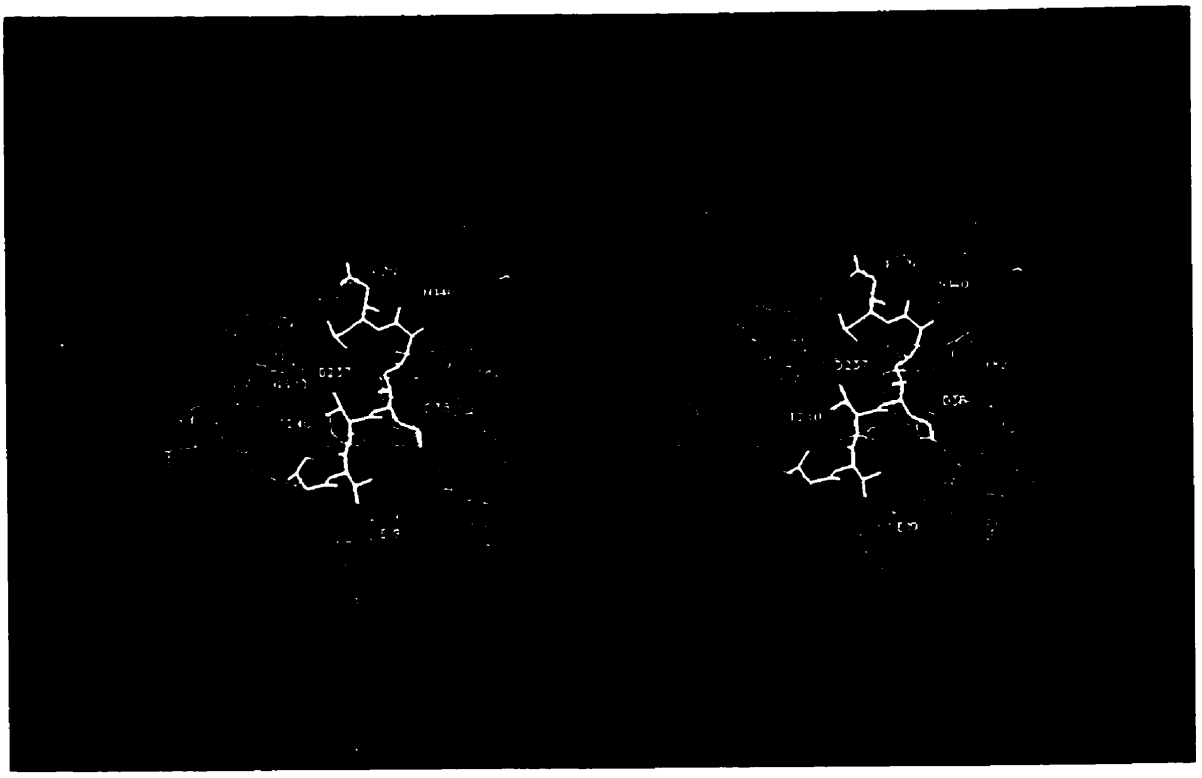


Figure 4.11. A diagram showing the substrate-binding cleft residues. Pepstatin A is shown in white. This diagram is generated by SETOR (Evans, 1993).

P3'-P4', which should interact with subsite S3', binds back into subsite S1' and forms an inverse γ-turn in the inhibitor. Due to the movement of the leucyl side chain, the carboxyl group of this statine residue is also away from the subsite S4' and is pointing towards the solvent. In the complex, the active-site flap (residues 82-88) undergoes a significant conformational change. It moves close to the inhibitor pepstatin A and covers up positions P2-P1'. Its

residues are involved in the formation of subsites S2, S1 and S2'. This flap becomes rigid due to the formation of hydrogen bonds and the van der Waals interactions with the inhibitor. Residue Tyr82 is involved in the formation of subsites S2' and S1, and separates these two subsites. At the S4 subsite, Leu321 moves about 0.5-0.7 Å toward the side chain of IVA to make good van der Waals contacts. Residue Ile244 is about 5 Å from IVA and forms part of the subsite S4. This suggests that modifications to a large group at the P4 position of the inhibitor may result in better contacts with the subsite and may confer stronger inhibition. The IVA side chain also interacts with the P2 valine residue by van der Waals contacts (d = 3.8 Å). At the subsite S3, the subsite residues make good interactions with the side chain of the valine residue at the P3 position of the inhibitor. Residue Leu17, which may be involved in the formation of the S3 subsite, is about 4-5 Å from the side chain of the valine residue, and therefore, there is room for a larger side chain. From the above analysis, it can be seen that pepstatin A has more hydrogen bonding and van der Waals interactions with the enzyme at the central region than at either end. Therefore, the inhibitor is tightly bound in the central region and loosely bound at both ends.

Table 4.4. van der Waals contacts between the enzyme and pepstatin A ($d \leq 4.0$ Å)

Pepstatin A	IVA	Val	Val	STA	Ala	STA
Position	P4	P3	P2	P1-P1'	P2'	P3'-P4'
Residues of	Leu321	Pro117	Gly325	Pro117	Gly40	Phe210
the enzyme	Phe242	Glu19	Thr240	Leu132	Phe210	Thr235
which make	Asn241	Asn241	Ile329	Tyr82	Asn140	Ile329
van der Waals		Thr240	Gly83	Asp38	Tyr82	Asp237
contacts with			Gly84	Asp237		
pepstatin A				Leu36		
				Gly40		
				Gly239		

4.2.7 Comparison of the pepstatin A conformations in different aspartic proteinases.

Pepstatin A is a universal inhibitor for all aspartic proteinases. Therefore, it is important to know whether pepstatin A is in the same conformation in all aspartic proteinase-pepstatin A complexes. The Protein Data Bank at the Brookhaven National Laboratory, USA, was searched and four aspartic proteinase-pepstatin A complex structures were found: endothiapepsin (Bailey *et al.*, 1993), rhizopuspepsin (Suguna *et al.*, 1992), human pepsin 3A (Fujinaga *et al.*, 1995) and human cathepsin D (Baldwin *et al.*, 1993). The pepstatin A molecules in these four complexes were taken out and superimposed with the pepstatin A molecule in the RMP complex. The superimposition is shown in Figure 4.12. The pepstatin A conformation is more conserved in the central region (P3 to P2') than at both ends. Selected dihedral angles at the central region are listed in Table 4.5. The leucyl-like side chain of the statine residue at the P3'-P4' position binds into subsite S1' in all five complexes. However, the γ-turn is formed only in the endothiapepsin, RMP and rhizopuspepsin complexes. The O...N distances and $\angle$O...H-N angles in these five pepstatin A structures are summarized in Table 4.6. Because of the reverse of the main chain at the P3' position for all

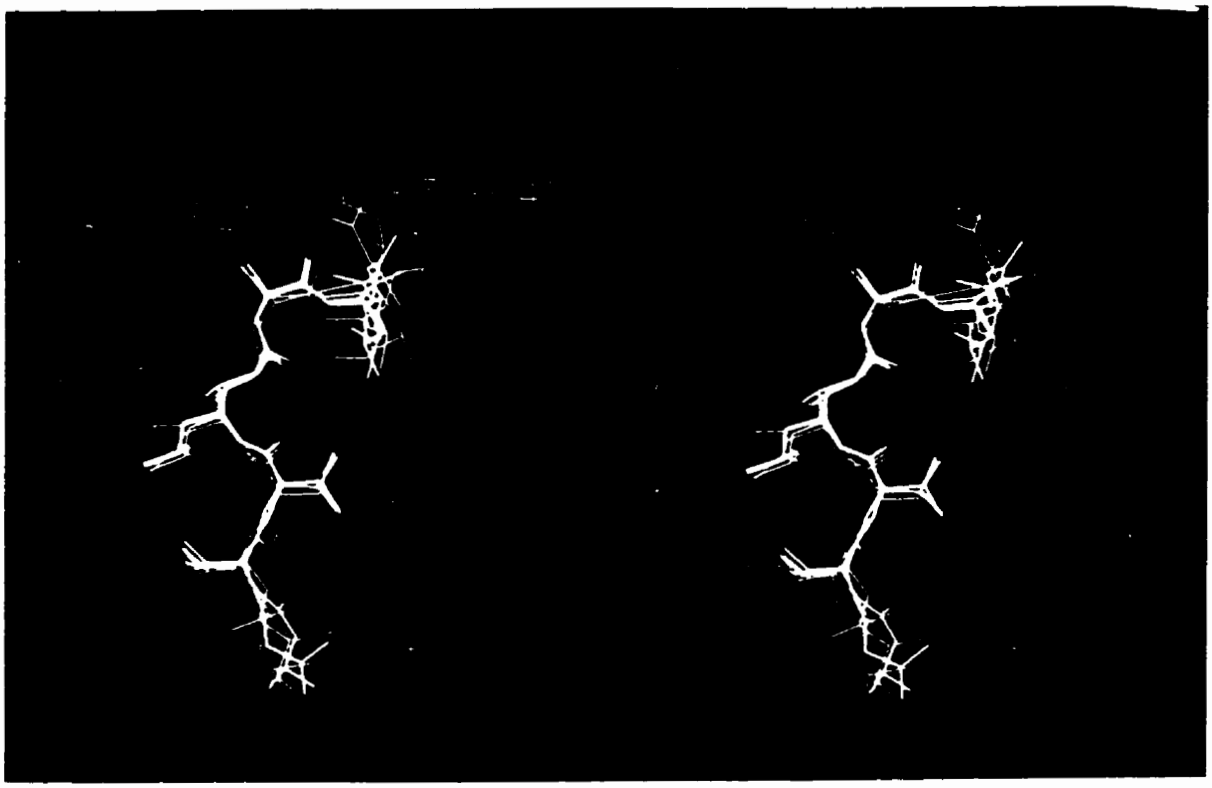


Figure 4.12. The superposition of the pepstatin A molecules from the RMP (green), endothiapepsin (white), rhizopuspepsin (yellow), human cathepsin D (pink) and human pepsin 3A (blue) molecules complexed with pepstatin A.

five complexes, the carboxyl group at the P4' position of the P3'-P4' statine residue is away from the subsite S4' and is in a different conformation for each complex. The P1-P2' part of the main chain in the pepstatin A molecule in the RMP-pepstatin A complex is about 0.5 to 1.0 Å away from the mean of this part of the other four conformations. The hydroxyl group at the C3 position of the P1-P1' statine residue in the RMP complex is in a different

Table 4.5. Selected dihedral angles in the central region of the pepstatin A molecule in the different pepstatin A-aspartic ptoteinase complexes.

	N_3-C_{A3}-C_3-N_4	C_{A3}-C_3-N_4-C_{A4}	C_3-N_4-C_{A4}-C_{H4}	N_4-C_{A4}-C_{H4}-C_{M4}	C_{A4}-C_{H4}-C_{M4}-C_4	C_{H4}-C_{M4}-C_4-N_5
RMP	126.4°	-178.6°	-134.9°	62.7°	76.8°	91.1°
Endothiapepsin	99.8°	-175.2°	-123.4°	59.5°	85.4°	84.7°
Rhizopuspepsin	97.4°	-179.2°	-111.7°	70.3°	63.8°	91.5°
human pepsin 3A	98.9°	-173.5°	-120.4°	63.7°	75.6°	82.0°
Cathepsin D	115.9°	-173.6°	-126.5°	63.7°	78.6°	79.7°

The numbering used is IVA(1)-Val(2)-Val(3)-STA(4)-Ala(5)-STA(6). The α-carbon atoms in the residues of the pepstatin A molecule are labeled as CA and the naming scheme used for the two dipeptide analogue statine residues is:

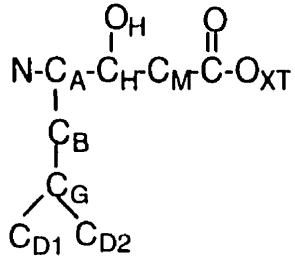


Table 4.6. The O...N distance and ∠O...H-N angle (P3'-P1') in different pepstatin A molecules in the different complex structures.

	O...N distance (Å)	∠O...H-N angle
RMP	2.86	124°
endothiapepsin	3.05	137°
rhizopuspepsin	3.17	135°
human pepsin 3A	3.84	76°
human cathepsin D	3.08	86°

conformation from the other hydroxyl group conformations, which are approximately the same in the other four pepstatin A molecules. The C_A-C_H-C_M-O plane of the statine residue at the P1-P1' position, which contains the hydroxyl group, is at about 50° from the mean plane of the other four conformations. In summary, the P3-P2' region of the pepstatin A molecule is more conserved in the inhibitor binding than both ends. Because of the differences of the residues at the substrate-binding site of the aspartic proteinases, the conformation of the P3-P2' region of the pepstatin A molecule undergoes minor conformational changes to provide better binding.

4.3 Conclusion

The crystal structure of the *Rhizomucor miehei* aspartic proteinase (RMP) complexed with the inhibitor pepstatin A has been refined to a crystallographic *R*-value of 19.3% and a *R*-free value of 28.0% at 2.7 Å. In the final model, a pepstatin A molecule fits into the large substrate-binding cleft between the two domains of RMP in an extended conformation up to the alanine residue at P2' position. The dipeptide analogue statine residue at the P3'-P4' position forms an inverse γ-turn (P3'–P1'), with its leucyl side chain binding into the S1' subsite. The inhibitor interacts with the residues of the substrate binding pocket by either hydrogen bonds or hydrophobic interactions, or both. The hydroxyl group of the statine residue at the P1-P1' position forms hydrogen bonds with both the catalytic aspartate residues (Asp38 and Asp237). This conformation mimics the expected transition state of the enzyme-substrate reaction. The binding of the inhibitor to the enzyme does not produce large distortions of the active site. No domain movement was observed in the complex compared to the native enzyme structure. However, the surface flap region (residues 82 to 88) undergoes a conformational change. It moves toward the inhibitor and becomes more rigid due to the formation of hydrogen bonds with the inhibitor. Opposite to the

surface flap at the active site, a surface loop (residues 323 to 328) also undergoes a conformation change. It moves about 1 to 1.5 Å towards the N-terminus of the inhibitor pepstatin A. Comparison of B-factors of the two domains suggests that the C-terminal domain becomes more rigid in the complex than it is in the native structure.

4.4 Future perspectives

The crystal structures of RMP and its pepstatin A complex have been determined at 2.15 Å and 2.7 Å, respectively. Because of the relatively low resolution of the complex structure, no solvent molecules were added to the structure. It would be appropriate to extend the RMP-pepstatin A complex structure to higher resolution. A high resolution structure would certainly provide a more accurate description of the active site in the complex. Water molecules should be located so that we could know whether any water molecules are conserved in the aspartic proteinase-pepstatin A complexes, and possibly determine their biological functions in the proteolysis process.

In section 3.2.5, it was proposed that the electrostatic potential at the active site determines the optimal pH of activity for each aspartic proteinases. The active site electrostatic potentials of the hypothetical mutants Asp303Ala chymosin b, Ala317Asp renin and Glu19Ala RMP were calculated. It is important to carry out the mutagenesis studies, determine the structures and measure the pH optima for these mutants to decide if the hypothesis can be validated.

In section 3.2.6, the relationship between attached carbohydrate moieties and the thermal stability of RMP was discussed. A new hypothesis to explain how carbohydrate moieties confer glycoproteins with high thermal stability was proposed. The content of β-structure in RMP decreased with the prolonged oxidation. Therefore, structural studies of this oxidized RMP molecule should be carried out to find out how the RMP molecule has changed due to oxidation, and to determine the reason for the decrease of the enzyme activity and the relation between the carbohydrate moieties and the enzyme activity. Mutagenesis studies of MPP, an RMP analogue, by Yamashita *et al.* (1994) show that substitution of Gly186 by various other amino acids decreases the thermal stability of MPP. The main chain torsion angles (ϕ, ψ) of Gly186 are in a region that is not allowed for the non-glycine residues in the Ramachandran plot. Therefore, it is appropriate to carry out mutagenesis studies of RMP and determine some of the structures of the mutants. From such studies, it should be possible to determine how this single residue Gly186 can cause such a large change in thermal stability and whether this residue can cause any conformational change of the carbohydrate moiety at Asn188.

From the structural comparison studies in Section 3.2.7, it is apparent that RMP and MPP diverged from other aspartic proteinases at an early stage

of evolution and form a sub-family of aspartic proteinases. The molecular evolution of aspartic proteinases should be studied, especially the evolution of the substrate binding site. HIV proteinases may develop resistance to the inhibitors designed thus far. Therefore, substrate-binding site evolution studies might provide a guideline for the synthesis of more potent HIV inhibitors. A potent inhibitor, which is resistant to HIV mutations, might be a great breakthrough in the treatment of AIDS.

References

IUPAC-IUB Commission on Biochemical Nomenclature (1972). A one-letter notation for amino acid sequences. *Pure Appl. Chem.* **31**, 641-645.

Abad-Zapatero, C., Rydel, T.J., Neidhart, D.J., Luly, J. & Erickson, J.W. (1991). Inhibitor binding induces structural changes in porcine pepsin. In *Advances in Experimental Medicine and Biology* (Bunn, B.M., ed.), vol. 306, pp. 9-21, Plenum Press, New York.

Aikawa, J., Yamashita, T., Nishiyama, M., Horinouchi, S. & Beppu T. (1990). Effects of glycosylation on the secretion and enzyme activity of *Mucor* renin, an aspartic proteinase of *Mucor pusillus*, produced by recombinant yeast. *J. Biol. Chem.* **265**, 13955-13959.

Andreeva, N.S. & James, M.N.G. (1991). Why does pepsin have a negative charge at very low pH? An analysis of conserved charged residues in aspartic proteinases. In *Advances in Experimental Medicine and Biology* (Bunn, B.M., ed.), vol. 306, pp. 39-45, Plenum Press, New York.

Aschaffenburg, R. (1968). Review of the progress of dairy science. Section G. Genetics. Genetic varients of milk proteins: Their breed distribution. *J. Dairy Res.* **35**, 447-460.

Bailey, D., Cooper, J.B., Veerapandian, B., Blundell, T.L., Atrash, B., Jones, D.M. & Szelke, M (1993). X-ray-crystallographic studies of complexes of pepstatin A and a statin-containing human renin inhibitor with endothiapepsin. *Biochem. J.* **289** (Pt 2), 363-371.

Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C. II, Cachau, R.E., Collins, J., Silva, A.M. & Erickson, J.W. (1993). Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. USA*, **90**, 6796-6800.

Baudyš, M., Foundling, S., Pavlik, M., Blundell, T. & Kostka, V. (1988). Protein chemical characterization of *Mucor pusillus* aspartic proteinase. Amino acid sequence homology with the other aspartic proteinases, disulfide bond arrangement and site of carbohydrate attachment. *FEBS Letters* **235**, 271-274.

Bech, A.M. & Foltmann, B. (1981). Partial primary structure of *Mucor miehei* protease. *Neth. Milk Dairy J.* **35**, 275-280.

Beintema, J.J. (1986). Do asparagine-linked carbohydrate chains in glycoproteins have a preference for beta-bends? *Biosci. Rep.* **6**, 709-714.

Betz, S.F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci.* **2**, 1551-1558.

Blundell, T.L. & Johnson, L.N. (1976). Protein crystallography. Academic Press Inc., New York.

Boel, E., Bech, A.M., Randrup, K., Draeger, B., Fiil, N.P. & Foltmann, B. (1986). Primary structure of a precursor to the aspartic proteinase from *Rhizomucor miehei* shows that the enzyme is synthesized as a zymogen. *Proteins Struct. Funct. Genet.* **1**, 363-369.

Boger, J., Lohr, N.S., Ulm, E.H., Poe, M., Blaine, E.H., Fanelli, G.M., Lin, T-Y, Payne, L.S., Schorn, T.W., LaMont, B.I., Vassil, T.C., Stabilito, I.I., Veber, D.F., Rich, D.H., & Bopari, A.S. (1983). Novel renin inhibitors containing the amino acid statine. *Nature* **303**, 81-84.

Bott, R., Subramanian, E. & Davies D.R. (1982). Three-dimensional structure of the complex of the *Rhizopus chinensis* carboxyl proteinase and pepstatin at 2.5 Å resolution. *Biochemistry* **21**, 6956-6962.

Brown, E.D. & Yada, R.Y. (1991). A kinetic and equilibrium study of the denaturation of aspartic proteinases from fungi, *Endothia parasitica* and *Mucor miehei. Biochim. Biophys. Acta* **1076**, 406-415.

Brown, R.J. & Ernstrom, C.A. (1988). Milk-clotting enzymes and cheese chemistry, part I: milk-clotting enzymes. In *Fundamentals of Dairy Chemistry* (Wong, N.P., Jenness, R., Keeney, M. & Marth, E.H., ed.), pp. 609-633, Van Nostrand Reinhold Company, New York, USA.

Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Brünger, A.T. (1992a). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **55**, 472-475.

Brünger, A.T. (1992b). In *X-PLOR Manual Version 3.1*, Yale University, New Haven, CT 06511, U.S.A.

Brünger, A.T. (1993). Assessment of phase accuracy by cross validation: the free *R* value. Methods and applications. *Acta Crystallog. sect. D*, **49**, 24-36.

Chen, H.M., Ford, C. & Reilly, P.J. (1994). Substitution of asparagine residues in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation. *Biochem. J.* **301**, 275-281.

Chothia, C. & Janin, J. (1982). Orthogonal packing of β-pleated sheets in proteins. *Biochemistry* **21**, 3955-3965.

Collaborative Computational Project, no. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.

Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K.E., Deutsch, P.J. & Emini, E.A. (1995). In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* **374**, 569-571.

Cooper, J.B., Foundling, S.I., Blundell T.L., Boger, J., Jupp, R.A., & Kay, J. (1989). X-ray studies of aspartic proteinase-statine inhibitor complexes. *Biochemistry* **28**, 8596-8603.

Cooper, J.B., Kahn, G., Taylor, G., Tickle, I.J. & Blundell, T.L. (1990). X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**, 199-222.

De Konning, P.J. & Draaisma, J. Th (1973). Identification of different types of rennet by means of isoelectric focusing. *Neth. Milk Dairy J.* **27**, 368-378.

De Vito, E., Guardia, D.C. & Martinez de Melián, E.R. (1997). Spontaneous inhibition or inactivation of renin in rat plasma. *Comp. Biochem. Physiol. C: Pharmacol. Toxicol. Endocrinol.* **116**, 55-60.

Devlin, T.M. (1982). Textbook of biochemistry with clinical correlations. John Wiley & Sons, Inc., New York.

Doan, D.N.P. & Fincher, G.B. (1992). Differences in the thermo-stabilities of barley (1→3, 1→4)-β-glucanases are only partly determined by N-glycosylation. *FEBS Letters* **309**, 265-271.

Drenth, J. (1994). Principles of protein X-ray crystallography. Springer-Verlag New York, Inc., New York.

Evans, S.V. (1993). SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134-138.

Farkye, N.Y. (1995). Contribution of milk-clotting enzymes and plasmin to cheese ripening. In *Advances in Experimental Medicine and Biology* (Malin, E.L. & Tunick, M.H., ed.), vol. 367, pp.195-207, Plenum Press, New York.

Fersht, A.R., Shi, J., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y. & Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235-238.

Foundling, S.I., Cooper, J., Watson, F.E., Cleasby, A., Pearl, L.H., Sibanda, B.L., Hemmings, A., Wood, S.P., Blundell, T.L., Valler, M.J., Norey, C.G., Kay, J., Boger, J., Dunn, B.M., Leckie, B.J., Jones, D.M., Atrash, B., Hallett, A., & Szelke, M. (1987a). High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. *Nature* **327**, 349-352.

Foundling, S.I., Cooper, J., Watson, F.E., Pearl, L.H., Hemmings, A., Wood, S.P., Blundell, T., Hallett, A., Jones, D.M., Suerias, J., Atrash, B., & Szelke, M. (1987b). Crystallographic studies of reduced bond inhibitors complexed with an aspartic proteinase. *J. Cardiovasc. Pharmacol.* **10** (*suppl.* **7**), .S59-S68.

Fox, P.F. (1993). Cheese: an overview. *In Cheese: Chemistry, Physics and Microbiology* (Fox, P.F., ed.), vol. I, pp. 1-36, Chapman and Hall, London.

Fujinaga, M., Chernaia, M.M., Tarasova, N.I., Mosimann, S.C. & James, M.N.G. (1995). Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci.* **4**, 960-972.

Glusker, J.P. & Trueblood, K.N. (1985). Crystal structure analysis: A primer. Oxford University Press, New York.

Goldman, A. (1995). How to make my blood boil. *Structure* **3**, 1277-1279.

Gray, G.L., Hayenga, K., Cullen, D., Wilson, L.J. & Norton, S. (1986). Primary structure of *Mucor miehei* aspartyl protease: evidence for a zymogen intermediate. *Gene* **48**, 41-53.

Harboe, M. (1996). *Rhizomucor miehei* aspartic proteinase having improved properties. In *Abstracts of VIIth International Aspartic Proteinase Conference*, M5-3, Banff, Canada.

Ho, D.D., Toyoshima, T., Mo, H., Kempf, D.J., Norbeck, D., Chen, C.M., Wideburg, N.E., Burt, S.K., Erickson, J.W. & Singh, M.K. (1994). Characterization of human immunodeficiency virus type 1 variants with increased resistance to a C2-symmetric protease inhibitor. *J. Virol.* **68**, 2016-2020.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Holsinger, V.H., Smith, P.W. & Tunick, M.H. (1995). Overview: cheese chemistry and rheology. In *Advances in Experimental Medicine and Biology* (Malin, E.L. & Tunick, M.H., ed.), vol. 367, pp.1-6, Plenum Press, New York.

Hoover, D.J., Veerapandian, B., Cooper, J.B., Damon, D.B., Dominy, B.W., Rosati, R.L., & Blundell T.L. (1991). X-ray analysis of a difluorostatone renin inhibitor bound as the tetrahedral hydrate to the aspartic protease endothiapepsin. In *Advances in Experimental Medicine and Biology* (Dunn, B.M., ed.), vol. 306, pp. 269-273, Plenum Press, New York.

Hyslop, D.B., Swanson, A.M. & Lund, D.B. (1979). Heat inactivation of milk-clotting enzymes at different pH. *J. Dairy Sci.* **62**, 1227-1232.

James, M.N.G., & Sielecki, A.R. (1985). Stereochemical analysis of peptide bond hydrolysis catalyzed by the aspartic proteinase penicillopepsin. *Biochemistry* **24**, 3701-3713.

James, M.N.G., Sielecki, A., Salituro, F., Rich, D.H. & Hofmann, T. (1982). Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin. *Proc. Natl. Acad. Sci.* USA, **79**, 6137-6141.

Jenness, R. (1988). Composition of milk. In *Fundamentals of Dairy Chemistry* (Wong, N.P., Jenness, R., Keeney, M. & Marth, E.H., ed.), pp. 1-38, Van Nostrand Reinhold Company, New York, USA.

Jia, Z., Vandonselaar, M., Schneider, P. & Quail, J.W. (1995). Crystallization and preliminary X-ray structure solution of *Rhizomucor miehei* aspartic proteinase. *Acta Crystallog. sect. D*, **51**, 243-244.

Johnson, A.H. (1974). The composition of milk. In Fundamentals of Dairy Chemistry (Webb, R.H., Johnson, A.H. & Alford, J.A., ed.), pp. 1-57, The AVI Publishing Company, Inc., Westport, Connecticut, USA.

Jossès, J., Schoentgen, F., Alais, C., Fiat, A. -M. & Jossès, P. (1972a). Studies on the primary structure of cow κ-casein. Structural features of para-κ-casein; N-terminal sequence of κ-casein-glycopeptides studied with a sequencer. *Helv. Chim. Acta* **55**, 2872-2883.

Jossès, J., Schoentgen, F., Alais, C. & Jossès, P. (1972b). Studies on the primary structure of cow κ-casein: The primary sequence of cow para-κ-casein. *Chimia* **26**, 645-646.

Kalan, E.B. & Woychik, J.H. (1965). Action of rennin on κ-casein, the amino acid composition of para-κ-casein. *J. Dairy Sci.* **48**, 1423-1428.

Kaplan, A.H., Michael, S.F., Wehbie, R.S., Knigge, M.F., Paul, D.A., Everitt, L., Kempf, D.J., Norbeck, D.W., Erickson, J.W. & Swanstrom, R. (1994). Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc. Natl. Acad. Sci. USA*, **91**, 5597-5601.

Kodama, S., Tsujimoto, M., Tsuruoka, N., Sugo, T., Endo, T. & Kobata A. (1993). Role of sugar chains in the in-vitro activity of recombinant human interleukin 5. *Eur. J. Biochem.* **211**, 903-908.

Korolev, S., Nayal, M., Barnes, W.M. & Waksman, G. (1995). Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5 Å resolution: structural basis for thermostability. *Proc. Natl. Acad. Sci.* USA, **92**, 9264-9268.

Lagrange, A., Paquet, D. & Alais, C. (1980). Comparative study of two *Mucor miehei* acid proteinases. Purification and some molecular properties. *Int. J. Biochem.* **11**, 347-352.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a new program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Ledward, D.A., Finch, A. & Rickert, W.S. (1975). Structural and functional determinants of *Mucor miehei* protease. V. Enthalpy changes upon thermal denaturation in solution of varying pH. *Biochim. Biophys. Acta* **379**, 426-30

Luzzati, P.V. (1952). Traitement statistique des erreurs dans la determination des structures cristalines. *Acta Crystallog.* **5**, 802-810.

Machin, P.A. (1985). Molecular replacement. *Proceedings of the Daresbury Study Weekend, 15-16 February* (Machin, P.A. ed.). SERC Daresbury Laboratory, Warrington, UK.

Mackinlay, A.G. & Wake, R.G. (1971). κ-Casein and its attack by rennin (chymosin). In *Milk Proteins* (McKenzie, H.A., ed.), vol. II, pp. 175-215, Academic Press, New York.

Mantafounis, D. & Pitts, J. (1990). Protein engineering of chymosin: modification of the optimum pH of enzyme catalysis. *Prot. Eng.* **3**, 605-609.

Markowitz, M. (1996). Protease inhibitors. What they are, how they work and when to use them. International Association of Physicians in AIDS Care, Medical Publications Corporation, Chicago.

Markowitz, M., Mo, H., Kempf, D.J., Norbeck, D.W., Bhat, T.N., Erickson, J.W. & Ho, D.D. (1995). Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor. *J. Virol.* **69**, 701-706.

Martin, P., Raymond, M.N., Bricas, E. & Dumas B.R. (1980). Kinetic studies on the action of *Mucor pusillus*, *Mucor miehei* acid proteases and Chymosins A and B on a synthetic chromophoric hexapeptide. *Biochim. Biophys. Acta* **612**, 410-420.

Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

McBride-Warren, P.A. & Rickert, W.S. (1973). Structural and functional determinants of *Mucor miehei* protease. II. Circular dichroic studies on the native and periodate-oxidized enzyme. *Biochim. Biophys. Acta* **328**, 52-60.

Mercier, A. -C., Brignon, G. & Ribadeau-Dumas, B. (1973). Primary structure of bovine κ-casein B. Complete sequence. *Eur. J. Biochem.* **35**, 222-235.

Miller, R.P., Poper, C.H., Wilson, C.W. & DeVito, E. (1972). Renin inhibition by pepstatin. *Biochem. Pharmacol.* **21**, 2941-2944.

Milner-White, E., Ross, B.M., Ismail, R., Belhadj-Mostefa, K. & Poet, R. (1988). One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins. *J. Mol. Biol.* **204**, 777-782.

Montreuil, J. (1984). Spatial structures of glycan chains of glycoproteins in relation to metabolism and function: survey of a decade of research. *Pure Appl. Chem.* **56**, 859-877.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157-163.

Newman, M., Watson, F., Roychowdhury, P., Jones, H., Badasso, M., Cleasby, A., Wood, S.P., Tickle, I.J. & Blundell, T.L. (1993). X-ray analysis of aspartic proteinases. V. Structure and refinement at 2.0 Å resolution of the aspartic proteinase from *Mucor pusillus*. *J. Mol. Biol.* **230**, 260-283.

Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

O, K., Hill, J.S. & Pritchard P.H. (1995). Role of N-linked glycosylation of lecithin:cholesterol acyltransferase in lipoprotein substrate specificity. *Biochim. Biophys. Acta* **1254**, 193-197.

Otto, M.J., Garber, S., Winslow, D.L., Reid, C.D., Aldrich, P., Jadhav, P.K., Patterson, C.E., Hodge, C.N. & Cheng, Y.S. (1993). In vitro isolation and identification of human immunodeficiency virus (HIV) variants with reduced sensitivity to C-2 symmetrical inhibitors of HIV type 1 protease. *Proc. Natl. Acad. Sci. USA*, **90**, 7543-7547.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Data collection and processing: Proceedings of the CCP4 study weekend, 29-30 January, 1993.* (Sawyer, L., Issacs, N. & Bailey S. eds.), pp. 56-62, Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK.

Pearl, L.H. (1987). The catalytic mechanism of aspartic proteinases. *FEBS Letters* **214**, 8-12.

Pitts, J.E., Crawford, M.D., Nugent, P.G., Wester, R.T., Cooper, J.B., Mantyla, A., Fagerstrom, R., & Nevalainen, H. (1995). The three-dimensional X-ray crystal structure of the aspartic proteinase native to *Trichoderma Reesei* complexed with a renin inhibitor CP-80794. In *Advances in Experimental Medicine and* Biology (Takashashi, K., ed.), vol. 362, pp. 543-547, Plenum Press, New York.

Powell, M.J.D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming* **12**, 241-254.

Prins, J. & Nielsen, T.K. (1970). Microbial rennet. *Mucor miehei. Process Biochem.* **5**, 34-35.

Ramachandran, G.N., & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283-437.

Read, R.J. (1986). Improved Fourier coefficents for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140-149.

Rich, D.H., Sun, E.T.O., & Ulm, E. (1980). Synthesis of analogues of the carboxyl protease inhibitor pepstatin. Effects of structure on inhibition of pepsin and renin. *J. Med. Chem.* **23**, 27-33.

Rickert, W.S. & McBride-Warren, P.A. (1974). Structural and functional determinants of *Mucor miehei* protease. III. Isolation and composition of the carbohydrate moiety. *Biochim. Biophys. Acta* **336**, 437-444.

Rickert, W.S. & McBride-Warren, P.A. (1977). Structural and functional determinants of *Mucor miehei* protease. IV. Inactivation of the enzyme by diazoacetyl norleucine methyl ester, pepstatin and 1,2-epoxy-3-(*p*-nitrophenoxy) propane. *Biochim. Biophys. Acta* **480**, 262-274.

Rosenthal, I. (1991). Milk and dairy products: properties and processing. VCH Verlagsgesellschaft mbH, Weinheim, Germany.

Rossmann, M.G. & Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta. Crystallog.* **15**, 24-31.

Roussel, A., Foutecilla-Camps, J.C. & Cambillau, C. (1990). TURBO-FRODO: A new program for protein crystallography and modelling. *Acta Crystallog. sect. A*, **46**, C66.

Saheki, T. & Holzer, H. (1975). Proteolytic activities in yeast. *Biochim. Biophys. Acta* **384**, 203-214.

Šali, A., Veerapandian, B., Cooper, J. B., Foundling, S.I., Hoover, D.J. & Blundell, T.L. (1989). High resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: the analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J.* **8**, 2179-2188.

Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Comm.* **27**, 157-162.

Sielecki, A. R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S. & James, M.N.G. (1990). Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J.Mol. Biol.* **214**, 143-170.

Sharp, K.A. & Nicholls, A. (1989). In *DelPhi V3.0 Manual.* Department of Biochemistry and Molecular Biophysics, Columbia University, USA.

Steele, J.L. (1995). Contribution of lactic acid bacteria to cheese ripening. In *Advances in Experimental Medicine and Biology* (Malin, E.L. & Tunick, M.H., ed.), vol. 367, pp.209-220, Plenum Press, New York.

Sternberg, M. Z. (1971). Crystalline milk clotting protease from *Mucor miehei* and some of its properties. *J. Dairy Sci.* **54**, 159-167.

Sternberg, M. (1972). Bond specificity, active site and milk clotting mechanism of the *Mucor miehei* protease. *Biochim. Biophys. Acta* **285**, 383-392.

Stout, G.H. & Jensen, L.H. (1989). X-ray structure determination: A practical guide. John Wiley & Sons, Inc.,New York.

Suguna, K., Padlan, E.A., Bott, R., Boger, J., Parris, K.D., & Davies, D.R. (1992). Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins* **13**, 195-205.

Suguna, K., Padlan, E.A., Smith, C.W., Carlson, W.D. & Davies, D.R. (1987). Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action. *Proc. Nat. Acad. Sci., U.S.A.* **34**, 7009-7013.

Swaisgood, H.E. (1975). Methods of gel electrophoresis of milk proteins. American Dairy Science Association, Champaign, USA.

Szecsi, P.B. (1992). The aspartic proteinases. *Scand. J. Clin. Lab. Invest. Suppl.* **210**, 5-22.

Tang, J., James, M.N.G., Hsu, I.N., Jenkins, J.A. & Blundell, T.L. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**, 618-621.

Terashima, M., Kubo, A., Suzawa, M., Itoh, Y. & Katoh, S. (1994). The roles of the N-linked carbohydrate chain of rice α-amylase in thermo-stabilities and enzyme kinetics. *Eur. J. Biochem.* **226**, 249-254.

Thaller, C., Weaver, L.H., Eichele, G., Wilson, E., Karlsson, R. & Jansonius, J.N. (1981). Repeated seeding technique for growing large single crystals of proteins. *J. Mol. Biol.* **147**, 465-469.

Thompson, M.P. (1970). Phenotyping milk proteins: A review. *J. Dairy Sci.* **53**, 1341-1348.

Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M. & Hamada, M. (1970). Pepstatin, a new pepsin inhibitor produced by *Actinomycetes*. *J. Antibiot. Tokyo* **23**, 259-262.

Veerapandian, B., Cooper, J.B., Šali, A., Blundell, T.L., Rosati, R.L., Dominy, B.W., Damon, D.B., & Hoover, D.J. (1992). Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Science* **1**, 322-328.

Walstra, P. & Jenness, R. (1984). Dairy chemistry and physics. John Wiley & Sons, Inc., New York.

Wang, F.F.C. and Hirs, C.H.W. (1977). Influence of the heterosaccharides in porcine pancreatic ribonuclease on the conformation and stability of the protein. *J. Biol. Chem.* **252**, 8358-8364.

Wasserman, B.P. and Hultin, H.O. (1981). Effect of deglycosylation on the stability of *Aspergillus niger* catalase fungi. *Arch. Biochem. Biophys.* **212**, 385-392.

Waugh, D.F. & Von Hippel, P.H. (1956). κ-casein and the stabilization of casein micelles. *J. Am. Chem. Soc.* **78**, 4576-4582.

Whitney, R. McL., Brunner, J.R., Ebner, K.E., Farrel, H.M., Jr., Josephson, R.V., Morr, C.V. & Swaisgood, H.E. (1976). Nomenclature of proteins of cows' milk: Fouth revision. *J. Dairy Sci.* **59**, 785-815.

Yada, R.Y. & Nakai, S. (1986a). Use of principal component analysis to study the relationship between physical/chemical properties and the milk-clotting to proteolytic activity ratio of some aspartyl proteinases. *J. Agric. Food Chem.* **34**, 675-679.

Yada, R.Y. & Nakai, S. (1986b). Secondary structure of some aspartyl proteinases. *J. Food Biochem.* **10**, 155-183.

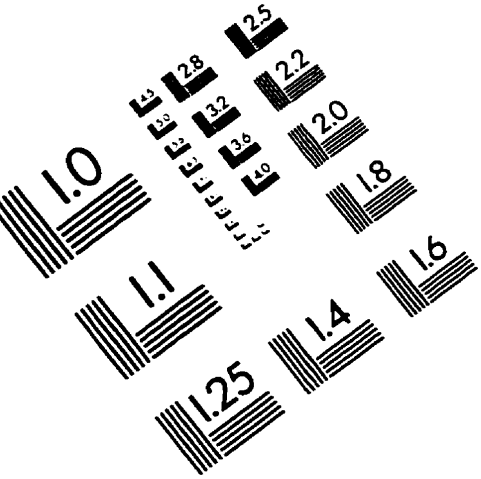
Yamashita, T., Higashi, S., Higashi, T., Machida, H., Iwasaki, S., Nishiyama, M. & Beppu, T. (1994). Mutation of a fungal aspartic proteinase, *Mucor pusillus* renin, to decrease thermostability for use as a milk coagulant. *J. Biotech.* **32**, 17-28.

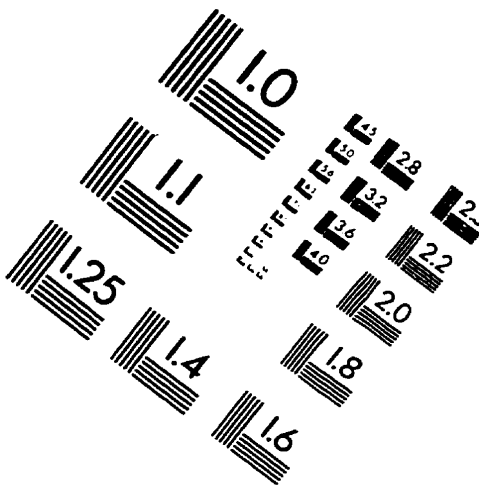
Yamauchi, T., Nagahama, M., Hori, H. & Murakami, K. (1988). Functional characterization of Asp-317 mutant of human renin expressed in COS cells. *FEBS Letters* **230**, 205-208.

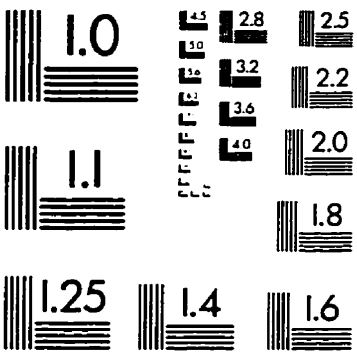
Yang, J., Teplyakov, A. & Quail, J.W. (1997). Crystal structure of the aspartic proteinase from *Rhizomucor miehei* at 2.15 Å resolution. *J. Mol. Biol.* **268**, 449-459.

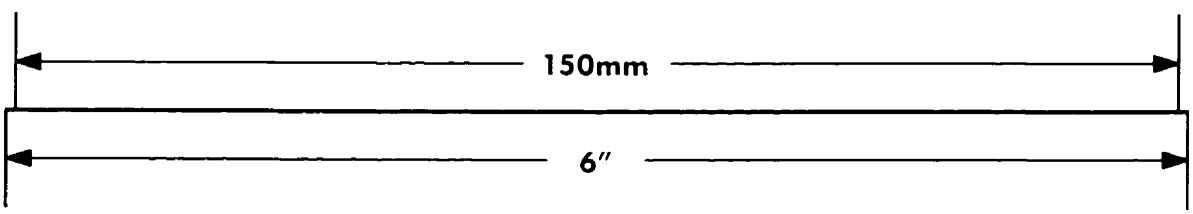
Yu, M.H., Weissman, J.S. & Kim, P.S. (1995). Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis. *J. Mol. Biol.* **249**, 388-397.

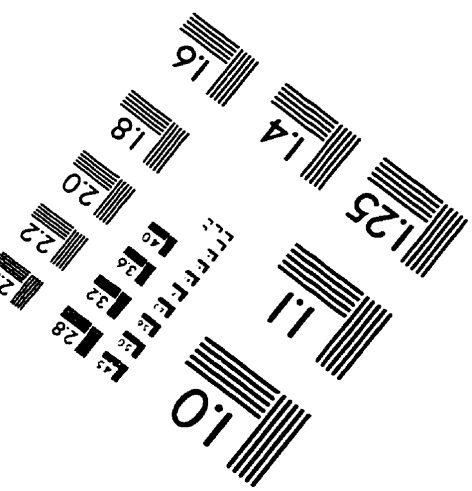
IMAGE EVALUATION TEST TARGET (QA-3)

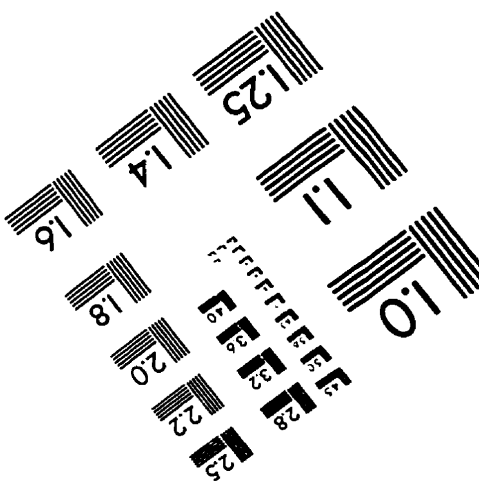












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989